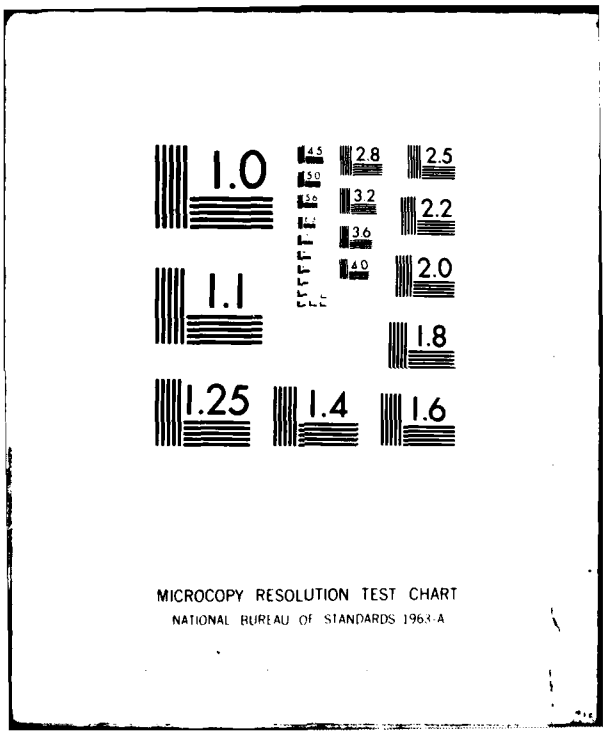
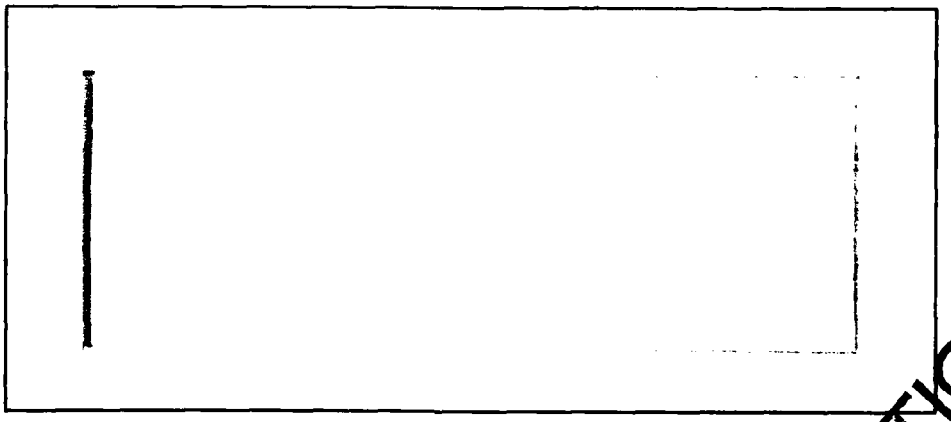
AD-A115 432 YALE UNIV NEW HAVEN CT DEPT OF COMPUTER SCIENCE F/G 12/1
LOCAL-MESH, LOCAL-ORDER, ADAPTIVE FINITE ELEMENT METHODS WITH A--ETC(U)
DEC 81 A WEISER N00014-76-C-0277
UNCLASSIFIED TR-213 NL


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963-A

(11)

AD A115432



DTIC FILE COPY

YALE UNIVERSITY
DEPARTMENT OF COMPUTER SCIENCE

Local-Mesh, Local-Order, Adaptive Finite Element Methods with A Posteriori Error Estimators for Elliptic Partial Differential Equations

Alan Weiser[1]

Technical Report #213

December, 1981

[1]Department of Computer Science, Yale University, New Haven, CT 06520

This research was supported in part by ONR Grant N00014-76-C-0277, FCU-ONR Grant F1.N00014-80-C-0076, the Department of Energy (DOE Grant DE-AC02-77ET53053), The National Science Foundation (Graduate Fellowship), and Yale University.

ABSTRACT

Local-Mesh, Local-Order, Adaptive Finite Element Methods with A Posteriori Error Estimators for Elliptic Partial Differential Equations

Alan Weiser

Yale University, 1981

The traditional error estimates for the finite element solution of elliptic partial differential equations are a priori, and little information is available from them about the actual error in a specific approximation to the solution. Thus, in the traditional finite element method, the choice of discretization is left to the user, who must base his decision on experience with similar equations.

In recent years, locally-computable a posteriori error estimators have been developed, which apply to the actual errors committed by the finite element method for a given discretization. These estimators lead to algorithms in which the computer itself adaptively decides how and when to generate discretizations. So far, for two-dimensional problems, the computer-generated discretizations have tended to use either local mesh refinement, or local order refinement, but not both.

In this thesis, we present a new class of local-mesh, local-order, square finite elements which can easily accommodate computer-chosen

→ next page

discretizations. We present several new locally-computable a posteriori error estimators which, under reasonable assumptions, asymptotically yield upper bounds on the actual errors committed, and algorithms in which the computer uses the error estimators to adaptively produce sequences of local-mesh, local-order discretizations. We present theoretical and numerical results indicating the expected and actual behavior of our methods.

Table of Contents

List of Figures

List of Tables

List of Symbols

CHAPTER 1

Introduction

The traditional error estimates for the finite element solution of elliptic partial differential equations are a priori, and little information is available from them about the actual error in a specific approximation to the solution. Thus, in the traditional finite element method, the choice of discretization is left to the user, who must base his decision on experience with similar equations.

In recent years, locally-computable a posteriori error estimators (e.g., Babuška and Rheinboldt [8]) have been developed, which apply to the actual errors committed by the finite element method for a given discretization. These estimators lead to algorithms in which the computer itself adaptively decides how and when to generate discretizations. So far, for two-dimensional problems, the computer-generated discretizations have tended to use either local mesh refinement (e.g., Babuška and Rheinboldt [10], Bank and Sherman [13]), or local order refinement (Babuška, Szabo, and Katz [11]), but not both.

In this thesis, we present a new class of local-mesh, local-order,

square finite elements which can easily accommodate computer-chosen discretizations. We present several new locally-computable a posteriori error estimators which, under reasonable assumptions, asymptotically yield upper bounds on the actual errors committed, and algorithms in which the computer uses the error estimators to adaptively produce sequences of local-mesh, local-order discretizations. We present theoretical and numerical results indicating the expected and actual behavior of our methods.

We begin in Chapter 2 by considering bisection-type local mesh refinement with square elements. We present the class of "1-irregular" meshes, a variant of a class of meshes suggested by Bank and Sherman [14], such that any bisection-type locally-refined mesh can be converted into a 1-irregular mesh with no more than a constant times as many elements, and the resulting finite element matrix (with bilinear basis functions) has no more than a constant number of nonzeroes in any row. Conversely, we show that there are simple bisection-type locally-refined meshes with essentially dense finite element matrices. We discuss simple data structures for handling 1-irregular meshes.

In Chapter 3, we extend the finite element spaces considered in Chapter 2 to include basis functions with locally-varying polynomial orders.

In Chapter 4, we develop new a posteriori error estimators. We present a new error estimator which has two main advantages over the

estimator presented in Babuška and Rheinboldt [8]:

1. Under reasonable assumptions, the new error estimator is asymptotically an upper bound on the norm of the true error.

2. The new error estimator can be computed locally in each element. (The error estimator in [8] must be computed over more complicated local regions.)

Under suitable assumptions, we show that, like the estimator in [8], the new error estimator is no larger than a constant times the norm of the actual error. We also present several cheaper error estimators, and discuss their properties. Under suitable assumptions, Babuška and Miller [4] have recently shown that the error estimator used in their code (with piecewise bilinear basis functions) converges to the norm of the true error. We have not yet proven convergence for our error estimators.

In Chapter 5, we develop several heuristic models of error behavior which lead to adaptive refinement procedures (cf. Brandt [16], Babuška and Rheinboldt [7]). We present the asymptotic expected error and work behavior for three problem classes consisting of problems with:

1. smooth solutions.
2. solutions with point singularities.
3. solutions with line singularities.

Except in the presence of line singularities, the expected error converges to zero faster than inverse polynomially with respect to work.

In Chapter 6, we discuss some of the computational aspects of the methods presented in Chapters 2-5. We consider efficient assembly of the finite element system (cf. Eisenstat and Schultz [18], Weiser, Eisenstat, and Schultz [31]), and efficient construction of the error estimators. We present the asymptotic complexity of nested-dissection-type Gaussian elimination for the three classes of problems considered in Chapter 5. The operation counts and storage for the problems with singularities are linear (optimal-order) in the number of elements N, in contrast to the operation counts and storage for uniform meshes, which are proportional to $N^{3/2}$ and N log(N) respectively (e.g., George [23]).

In Chapter 7, we present numerical results obtained using prototype codes. We present the resulting error and error estimator behavior for problems from the three classes considered in Chapter 5. The error estimators are usually accurate to within a factor of two. In some cases, the error estimators appear to converge to the norm of the true error as the mesh size decreases.

I am deeply grateful to my advisor, Professor Randolph E. Bank, for suggesting the topic of this research and working closely with me on its realization. I am honored to know him and to have worked with him. I would like to thank my other advisors, Professors Martin H. Schultz and Stanley C. Eisenstat, for helping to guide my research, and for their unfailing intelligence, integrity, and support.

Many other past and present members of the numerical analysis group

at Yale have contributed much to my enjoyment of graduate school and my appreciation and understanding of numerical analysis. They are (in order of appearance at Yale) Andy Sherman, John W. Lewis, Rati Chandra, Rob Schreiber, Jack Perry, Dave Fyfe, Trond Steihaug, Craig Douglas, Howie Elman, Ken Jackson, Tony Chan, John Kindle, and Benren Zhu.

I am grateful to many other members-at-large in the numerical analysis community. I am especially grateful to Professor Mary F. Wheeler, for her encouragement and guidance when I was an undergraduate, and Professor Ivo Babuška, for his excellent and influential work.

I am grateful to the Office of Naval Research (ONR Grant N00014-76-C-0277, FSU-ONR Grant F1.N00014-80-C-0076), the Department of Energy (DOE Grant DE-AC02-77ET53053), the National Science Foundation (Graduate Fellowship), and Yale University for their financial support.

Most of all, I am grateful to my family, and my loving wife Mary Ann.

CHAPTER 2

Local mesh refinement

2.1 Introduction

In this chapter, we consider bisection-type local mesh refinement with square elements. A sample mesh of this type is shown in Figure 2-1.

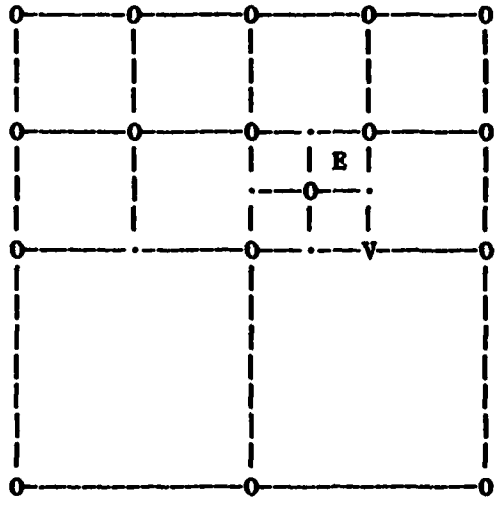


Figure 2-1: A bisection-type locally-refined mesh

This type of mesh refinement has been intensively investigated in recent years, mainly by Babuška and Rheinboldt and their students

(e.g., [4]). The data structures and computational algorithms required to implement the finite element method with general meshes of this type are fairly complicated (see Rheinboldt and Mesztenyi [25] and Gannon [21]). Thus, Bank and Sherman [14] and Van Rosendale [30] suggest using restricted classes of meshes which can be implemented with much simpler data structures.

In Section 2.2, we present some standard terminology for dealing with bisection-type local mesh refinement, and define the class of "1-irregular" meshes, a variant of a class of meshes suggested by Bank and Sherman [14]. In Section 2.3, we consider a variant of a mesh refinement rule suggested by Bank and Sherman which can convert any bisection-type locally-refined mesh into a 1-irregular mesh with no more than a constant times as many elements, such that the resulting finite element matrix (with bilinear basis functions) has no more than a constant number of nonzeroes in any row. Conversely, in Section 2.4, we show that there are simple bisection-type locally-refined meshes such that the resulting finite element matrices are essentially dense. We also consider alternative mesh refinement rules. In Section 2.5, we present data structure details for handling 1-irregular meshes.

2.2 Terminology

We now consider bisection-type local mesh refinement with square elements. For simplicity, we take the domain to be the unit square,

$$\Omega := [0,1]\times[0,1] = \{(x,y): 0\leq x\leq 1,\ 0\leq y\leq 1\},$$

with boundary $\partial\Omega$. The square region $E = [x_E, x_E+h_E]X[y_E, y_E+h_E]$ is called an element. The elements

$$[x_E, x_E+h_E/2]X[y_E+h_E/2, y_E+h_E], \quad [x_E+h_E/2, x_E+h_E]X[y_E+h_E/2, y_E+h_E],$$

$$[x_E, x_E+h_E/2]X[y_E, y_E+h_E/2], \quad [x_E+h_E/2, x_E+h_E]X[y_E, y_E+h_E/2]$$

are called the sons of E, and E is called the father of its sons. A mesh M is defined to be an unordered set of elements. An element in M is called refined if its four sons are in M, and unrefined otherwise.

The class of bisection-type locally-refined meshes with square elements, or admissible meshes (Babuška and Rheinboldt [8, 10]), is recursively defined by the following two rules:

1. $\{\Omega\}$ is an admissible mesh.

2. If a mesh M is an admissible mesh, and E is an unrefined element in M, then the mesh {M, the sons of E} is an admissible mesh.

The level L of E is defined to be the number of generations between Ω and E. Thus, in an admissible mesh on Ω, $h_E = (1/2)^L$. A neighbor of E is defined to be another element at the same level as E that shares a side with E.

A corner of an unrefined element is called a vertex. A vertex which is a corner of each unrefined element it touches is called a

regular vertex. All other vertices are called irregular*. The irregularity index (Babuška and Rheinboldt [10]) of a mesh is defined to be the maximum number of irregular vertices on a side of an unrefined element. We call a mesh with irregularity index $\leq k$ a k-irregular mesh. A (regular) vertex on $\partial\Omega$ is called a boundary vertex. All other vertices are called interior.

In figures depicting meshes,

- O denotes a regular vertex,
- o denotes a vertex (regular or irregular),
- · denotes an irregular vertex,
- (v) at a vertex denotes function value.

For example, Figure 2-1 depicts an admissible 2-irregular mesh with 4 refined elements, 13 unrefined elements (one labelled E), 12 regular boundary vertices, 5 regular interior vertices, and 6 irregular vertices (one labelled V).

---

*This is the name used by Babuška and Rheinboldt [10]. Bank and Sherman [14] use the name "green vertex". Gannon [21] uses the name "inactive vertex".

Let M be an admissible mesh. A function g is called <u>piecewise bilinear</u> on M if, on any unrefined element E in M,

$$\begin{aligned} g(x,y) = \; & g(x_E,y_E)\,(x_E+h_E-x)(y_E+h_E-y) \,/\, h_E^2 \\ & + g(x_E+h_E,y_E)\,(x-x_E)(y_E+h_E-y) \,/\, h_E^2 \\ & + g(x_E,y_E+h_E)\,(x_E+h_E-x)(y-y_E) \,/\, h_E^2 \\ & + g(x_E+h_E,y_E+h_E)\,(x-x_E)(y-y_E) \,/\, h_E^2 \,. \end{aligned}$$

Let the <u>finite element space</u> S (= S(M)) be the space of continuous piecewise bilinear functions on M. Let the regular vertices for M be $V_1,\ldots,V_N$. Let $B_I(x,y)$ (= $B(V_I)(x,y)$) be the function in S which satisfies the Lagrange conditions

$$B_I(V_J) = \delta_{IJ} = \begin{cases} 1 & \text{if } I = J \\ 0 & \text{if } I \neq J, \end{cases}$$

for J=1,...,N. B_I is well-defined using induction on element level. Moreover, $\{B_I\}$ is a basis for S, since each B_I is in S, the B_I's are linearly independent, and each function in S is a linear combination of the B_I's.

The <u>support</u> of B_I, supp(B_I), is defined to be the union of all unrefined elements E such that B_I is not identically zero in E. In general, the support of B_I may be non-convex or non-simply-connected. Alternately, the standard tensor-product basis functions $\{\underline{B}_I\}$ may be

defined for S, so that each supp($\underline{B}(V)$) is a rectangle. Most of the considerations applying to $\{B_I\}$ also apply to $\{\underline{B}_I\}$. However, supp(B(V)) is always contained in supp($\underline{B}(V)$) (see Figure 2-2), and we know of no computational advantage for $\{\underline{B}_I\}$. Thus, we restrict our attention to $\{B_I\}$.

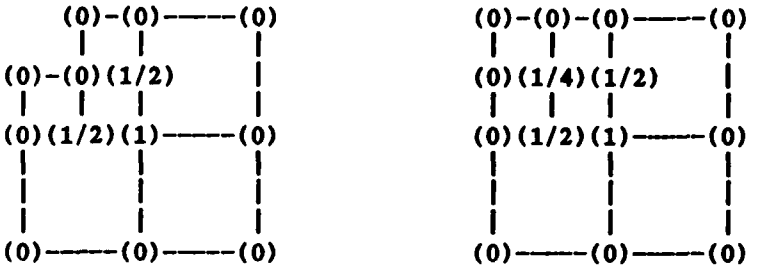


Figure 2-2: Lagrange versus tensor-product basis functions

2.3 The 1-irregular rule and some of its consequences

Given an admissible mesh, we force possible additional refinement by applying, as many times as possible, the following variant of a rule of Bank and Sherman [14]:

> refine any unrefined element with more than one irregular vertex on one of its sides. (2.1)

We call rule (2.1) the 1-irregular rule. The 1-irregular rule must terminate, since it never increases the maximum element level in the mesh. After the 1-irregular rule terminates, the resulting mesh is 1-irregular. Conversely, if any refinement forced by the 1-irregular

rule is omitted, the resulting mesh is not 1-irregular. Thus, the resulting mesh has the fewest elements of any 1-irregular mesh containing the original mesh.

The resulting configurations (up to symmetry) of the support of B(V) for interior V are depicted in Figure 2-3. Because of the 1-irregular rule, further refinement of an element in one of cases I-VI produces another of cases I-VI, possibly at the next finer level.

Each unrefined element E in the support of B(V) satisfies either

A : V is a corner vertex of E,

or

B : V is not a corner vertex of E, but
V is a corner vertex of the father of E, and
V is next to an irregular corner vertex of E.

Thus, it is simple to determine which basis functions are nonzero in E. Figure 2-4 depicts a sample E in its father, f(E), and numbers the 6 relevant vertices. B_3 and B_4 are nonzero in E (V_4 is regular since f(E) is refined). B_2 (B_1) is nonzero in E if V_2 is regular (irregular). B_5 (B_6) is nonzero in E if V_5 is regular (irregular).

There are exactly four basis functions nonzero in E. Since the element stiffness matrices are small (4X4), and the mappings from local to global basis functions are simple, the resulting finite element system is straightforward to assemble.

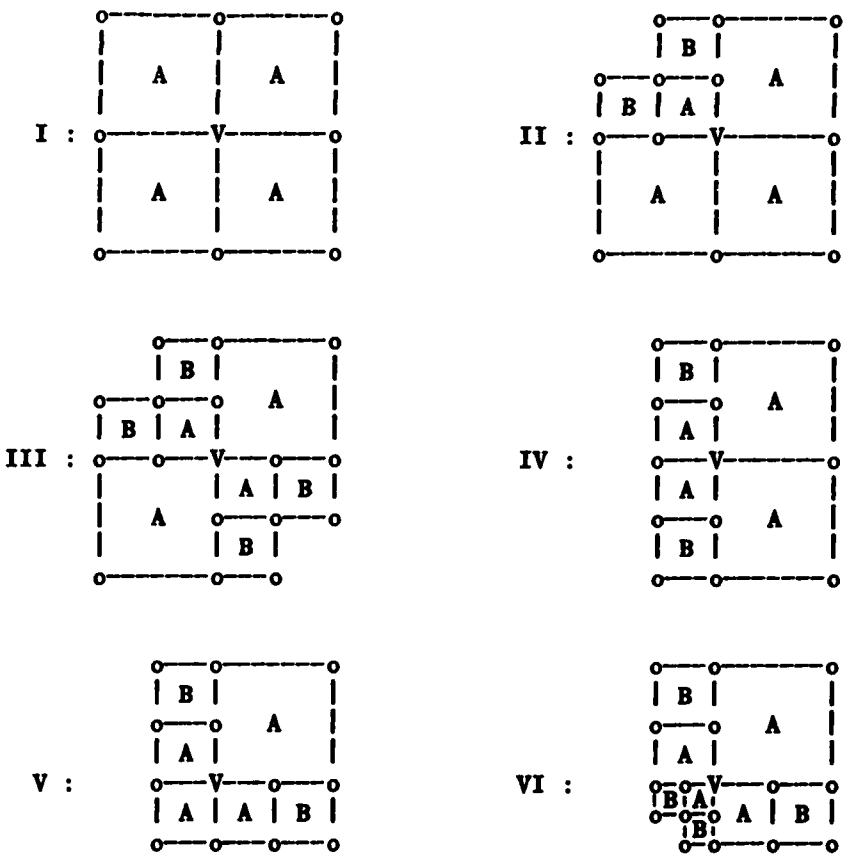


Figure 2-3: Configurations of elements in supp(B(V))

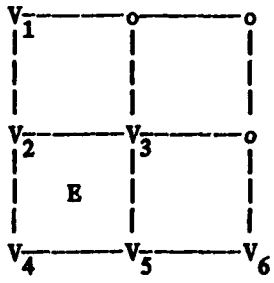


Figure 2-4: f(E)

Furthermore, the basis functions are locally independent:

in each unrefined element E, (2.2)

$\{B_I|_E : E \text{ in } \mathrm{supp}(B_I)\}$ are linearly independent,

where $\cdot|_E$ denotes restriction to element E. We can show that (2.2) holds if and only if

no irregular vertex touches unrefined elements (2.3)
at three different levels.

For example, in the mesh shown in Figure 2-1, irregular vertex V touches unrefined elements at levels 1, 2, and 3, and the restrictions to element E of the five basis functions nonzero in E are linearly dependent.

Let N(V) be the number of basis functions with support intersecting supp(B(V)), where V is a regular vertex in a 1-irregular mesh. N(V) is the number of nonzeroes in the row of the finite element matrix

corresponding to B(V). The minimum possible N(V) for an interior vertex is 5, and the maximum possible N(V) is 13 (Figure 2-5). N(V) is 9 for interior vertices in a locally uniform mesh. N(V) for boundary vertices can be as small as 4. N(V) is 6 for boundary side vertices in a locally uniform mesh.

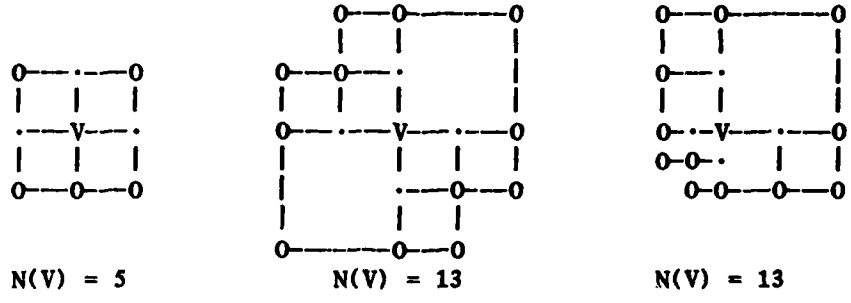


Figure 2-5: Minimum and maximum N(V) for interior vertices

Many refinement sequences arising in practice naturally yield 1-irregular meshes, so the 1-irregular rule introduces no additional refinement. We now show that, given an arbitrary admissible mesh, and applying the 1-irregular rule as many times as possible,

> the number of elements forced to refine by the 1-irregular rule is no more than eight times the number of refined elements in the original mesh. (2.4)

Hence the number of basis functions introduced by the 1-irregular rule is no more than a constant times the number of original basis functions.

Suppose there are LMAX levels in the mesh. Let

O(L) := {original refined elements at level L},

R(L) := {elements at level L refined by the 1-irregular rule}.

It suffices to show that each element E in R(L) shares a corner vertex with at least one element F in O(L). We do this using induction from L=LMAX back down to 0. There are clearly no elements in R(LMAX). Suppose the statement holds for all levels with index greater than L. Since E is in R(L), the 1-irregular rule forces E to refine because G, a neighbor of E, has a son, H, which is refined at level L+1, and shares a side with E. If H is in O(L+1), then G=F is in O(L), and (2.4) holds. Otherwise, H is in R(L+1). By induction, there is an element I in O(L+1) which shares a corner vertex with H. Then the father F of I is in O(L) and shares a corner vertex with E (Figure 2-6). So, (2.4) holds in all cases.*

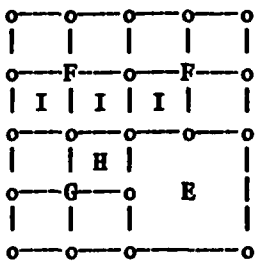


Figure 2-6: H in R(L+1)

---

*We do not know if (2.4) is sharp. The largest ratio of enforced refinement to original refinement we have encountered is 5.75.

In summary, given any admissible mesh, the 1-irregular rule forces additional refinement such that

the resulting 1-irregular mesh has no more than a constant times the original number of basis functions, (2.5)

and

the resulting element stiffness matrices are of bounded size (4X4), and simple to assemble, (2.6)

and

the number of nonzeroes in any row of the resulting finite element system is bounded independent of the mesh. (2.7)

2.4 Alternative mesh refinement rules

In this section, we show that a refinement rule like the 1-irregular rule is needed for properties (2.6) and (2.7). We also consider alternative mesh refinement rules.

Suppose, because of a sharp point singularity, that a mesh results from refining only elements which contain the point P = (2/3,2/3). Since

$$2/3 = 1 - 1/2 + 1/4 - 1/8 + 1/16 - \ldots ,$$

the mesh is formed by successively refining the alternately upper right or lower left son of the smallest refined element. If there are LMAX >> 1 levels in the mesh, the active vertices are the ten boundary vertices and the LMAX interior vertices

$$V_L = (P_L , P_L) , \quad P_L = \sum_{I=0}^{L} (-1/2)^I , \quad L=1,\ldots,LMAX.$$

Let element E be the smallest element containing P. For any V_L, it can be shown by induction on I that $B(V_L)$ is nonzero at the points (P_{I-1},P_I) and (P_I,P_{I-1}) if and only if $I > L$. Setting I = LMAX, since (P_{LMAX-1},P_{LMAX}) and (P_{LMAX},P_{LMAX-1}) are corners of E, E must be in the support of each $B(V_L)$. Thus, the element stiffness matrix for E, and the resulting finite element system, are essentially dense. Since the mesh is 2-irregular, this example disproves the conjecture (Babuška and Miller [4], page 17) that (2.7) holds for k-irregular meshes with k bounded.

Assembly of the finite element sytem is fairly complicated (e.g., Gannon [21], Rheinboldt and Mesztenyi [25]). If Gaussian elimination is used to solve the system, the solution time is asymptotically proportional to $LMAX^3$. By contrast, if the 1-irregular rule is applied, assembly is simple. If the vertices are ordered according to the maximum levels of the elements in their supports, solution time is asymptotically proportional to LMAX (see Section 6.3).

In the previous example, (2.6) and (2.7) can be insured without enforcing additional refinement, by using modified basis functions which are identically zero in the element containing the singularity. Figure 2-8 depicts the modified basis function $\underline{B}(V_1)$. When there are <u>two</u> sharp point singularities at

$$P_1 = (9/16 + (2/3)/16,\ 10/16 + (2/3)/16)$$

and

$$P_2 = (9/16 + (2/3)/16,\ 13/16 + (2/3)/16),$$

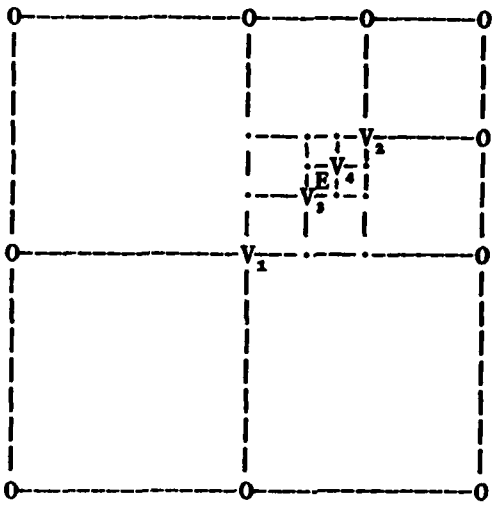


Figure 2-7: Singularity at P = (2/3,2/3), LMAX = 4

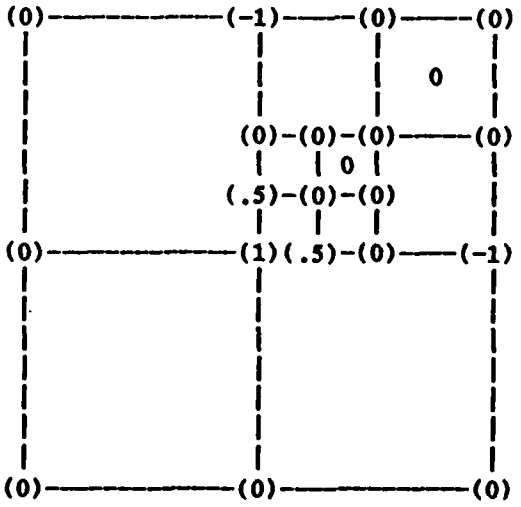


Figure 2-8: A modified basis function

however, the support of the basis function for V_1 must asymptotically intersect the supports of roughly half the basis functions in S, since

it can only be zero at one point along the line segment from (1/2,1/2) to (1/2,1). Thus, with <u>any</u> rule for forming basis functions, it may be advantageous (purely from matrix sparsity considerations) to force additional refinement.

We briefly mention alternatives to the 1-irregular rule. Another rule which insures properties (2.5), (2.6), and (2.7) (when applied as many times as possible) is:

> refine all unrefined side- and corner- neighbors of any element which has grandsons (sons of a son). (2.8)

Rule (2.8) is equivalent to a condition in Van Rosendale [30], page 59. With this rule, only cases I-V in Figure 2-3 are possible, and (2.5), (2.6), and (2.7) hold for both $\{B_I\}$ and $\{\underline{B}_I\}$. However, rule (2.8) may force more additional refinement, and hence require more work, than the 1-irregular rule, and is no easier to implement.

Conversely, any rule which may force less additional refinement than the 1-irregular rule does so only when neighboring elements with size ratios at least 4:1 occur. For example, if we apply the rule

> if an irregular vertex touches unrefined elements at three different levels, refine the element at the middle level. (2.9)

as many times as possible (refining elements with lowest level first), the resulting mesh will satisfy (2.6), and <u>may</u> have fewer elements than

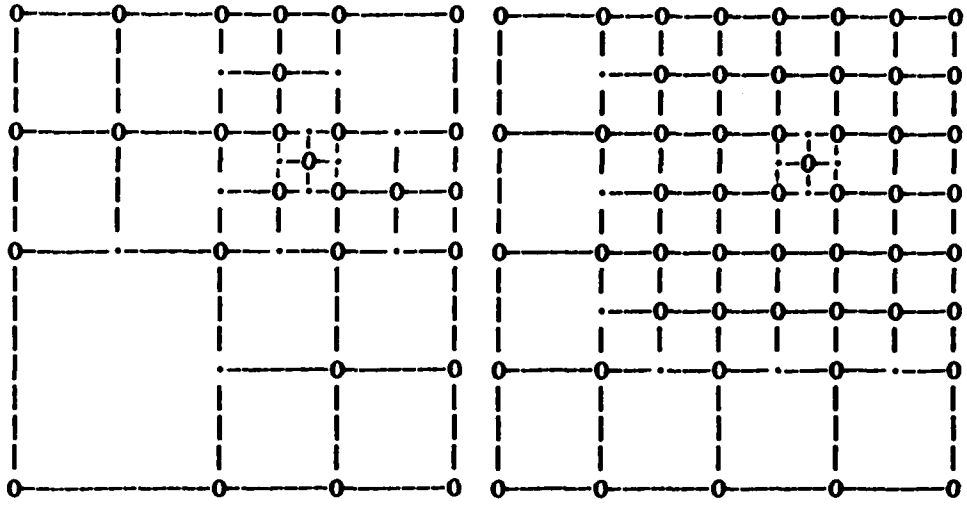
the resulting 1-irregular mesh. Unfortunately, (2.5) and (2.7) may not hold: for a problem with a sharp point singularity at (1/2+ε,1/2+ε), as ε approaches zero, the number of nonzeroes in the row of the finite element system for B(1/2,1/2) may be arbitrarily large.

Figure 2-9 depicts the results of applying the 1-irregular rule, rule (2.8), and rule (2.9) to the mesh in Figure 2-7.

2.5 A mesh data structure

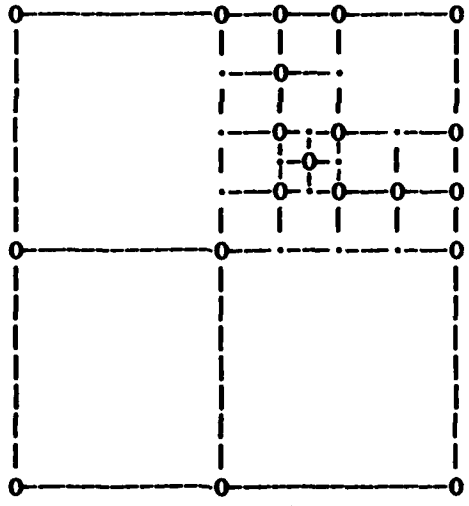
The data structure we use to represent 1-irregular meshes is a variant of a data structure presented in Bank and Sherman [14]. The data structure has two main components: a refined-element tree IRCT, stored in a 4X(NIRCT array, and a vertex list IVERT, stored in a 2X(NIVERT array. For each refined element E, there is a 4X4 node in IRCT containing pointers to the father of E, to any of the four sons of E which are refined, and to the nine vertices which are corners of the sons of E. For each interior vertex V which bisects an element side, there is a 2X1 node in IVERT containing pointers to the elements with sides bisected by V. Other types of vertices have other information stored in IVERT. Figure 2-10 depicts a sample refined element E, its node in IRCT, and a sample node in IVERT.

The neighbors of any element can be easily found. For instance, in Figure 2-10, the pointer location for the bottom neighbor of S_1 is in the node for E, while the pointer location for the right neighbor of S_1 is in the node for the right neighbor E' of E. The node for E' (if it



the 1-irregular rule

rule (2.8)



rule (2.9)

Figure 2-9: LMAX = 4: alternative mesh refinement rules

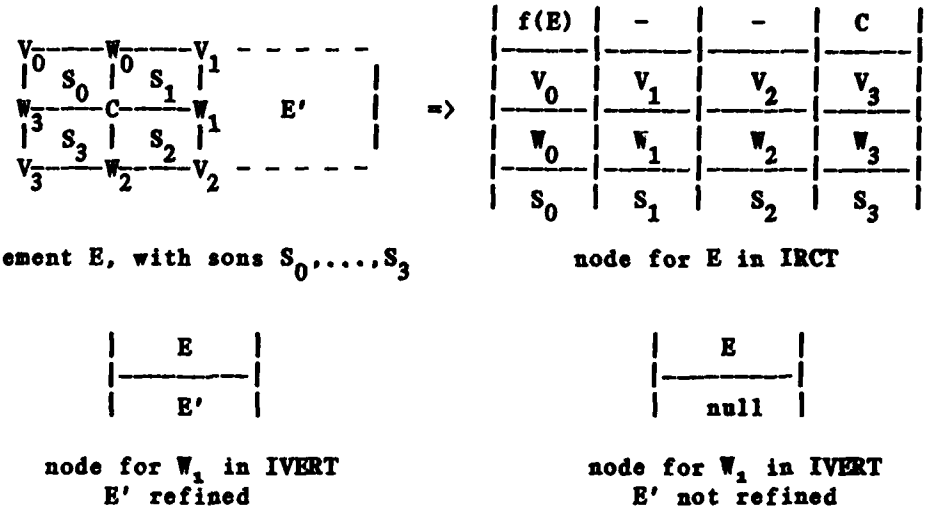


Figure 2-10: Nodes in IRCT and IVERT

exists) is pointed to in the node for W_1 in IVERT.

Moreover, suppose the entire mesh is rotated 90, 180, or 270 degrees. Then IVERT and the top row of IRCT remain the same, and the last three rows of each node in IRCT are rotated. For example, if the mesh is rotated 90 degrees counterclockwise, the node for E becomes

f(E)	-	-	C
V_1	V_2	V_3	V_0
W_1	W_2	W_3	W_0
S_1	S_2	S_3	S_0

.

Thus, the neighbors of any son of E can be found with the same procedure used to find the neighbors of S_1 , as long as indices into the last three rows of IRCT are incremented by the appropriate value mod 4.

This mesh data structure makes it relatively easy to construct the finite element system for S, to evaluate the approximate solution, and to refine the mesh. More data structure considerations are in Bank and Sherman [14].

CHAPTER 3

Local order refinement

3.1 Introduction

Traditional finite element methods for elliptic problems employ finite element spaces with low-order (e.g., bilinear) basis functions. These spaces are appropriate when the solutions to the elliptic problems are not smooth. When the solutions are smooth, standard approximation theory indicates that high-order piecewise polynomial basis functions can yield much faster convergence. Thus, it is advantageous to be able to use both low-order and high-order basis functions.

Local-order finite element spaces have received recent attention by Babuška, Katz, and Szabo [3, 11, 12] and Babuška and Dorr [2]. Full local-order, local-mesh methods in more than one dimension have not yet been considered.

In this chapter, we extend the finite element spaces with 1-irregular meshes considered in Chapter 2 to include basis functions with locally-varying polynomial orders. In Section 3.2, we consider the C^0 case, and in Section 3.3, we briefly consider spaces of smoother

functions. In Section 3.4, we present simple data structures for the spaces considered.

3.2 The Lagrange case

Let M be a given 1-irregular mesh. In this section, we construct a space S (= S(M)) of C^0 piecewise polynomials such that, in each unrefined element E in M,

smooth functions are locally approximated to order $k_E \geq 2$. (3.1)

Our local order refinement follows the basic approach of Taylor [29] and Babuška, Katz, and Szabo [3]. On [0,1], let the polynomials $\{p_i(x): i=1,\ldots,kmax\}$ satisfy

$p_1(x) = 1-x$,

$p_2(x) = x$,

and for $i > 2$,

$p_i(x)$ is a polynomial of order exactly i,

$p_i(0) = p_i(1) = 0$.

Since $\{p_1,\ldots,p_k\}$ is a basis for the polynomials of order $k > 1$, the p_i are called (C^0) hierarchic polynomials.

On element E, $p_i^E(x)$ and $p_m^E(y)$ are defined by mapping [0,1] onto $[x_E, x_E+h_E]$ and $[y_E, y_E+h_E]$:

$$D^j p_i^E(x) := D^j p_i((x-x_E)/h_E)\ h_E^{-j},$$

and

$$D^j p_m^E(y) := D^j p_m((y-y_E)/h_E)\ h_E^{-j},$$

$j=0,\ldots,kmax-1$.

Let $\bar{p}_i(t)$ be the Legendre polynomial on $[0,1]$ of order i. Since $\bar{p}_1(t) = 1$, and

$$\int_0^1 \bar{p}_i\ \bar{p}_j = 0\ ,\quad i \neq j,$$

a convenient choice for $\{p_i\}$ is

$$p_i(x) = \int_0^x \bar{p}_{i-1}(t)\ dt\ ,\quad i \geq 2. \tag{3.2}$$

We define S to be the span of three types of basis functions $\{B_I(x,y)\}$: <u>vertex</u>, <u>side</u>, and <u>element</u>.

The <u>vertex</u> basis functions are the piecewise bilinear basis functions from Chapter 2, with $B_I(V_J) = \delta_{IJ}$ for all regular vertices V_J.

The <u>side</u> basis functions are products of piecewise linear functions in one direction with quadratic or higher-order hierarchic polynomials in the other direction. Each side s of an unrefined element, not strictly contained in a side of another unrefined element, has side basis functions with hierarchic polynomials of order up to

$$k_s := \max\{k_E : s \text{ contains a side of unrefined element } E\}. \tag{3.3}$$

A basis function for side s in Figure 3-1 is $h(x)p_3^E(y)$.

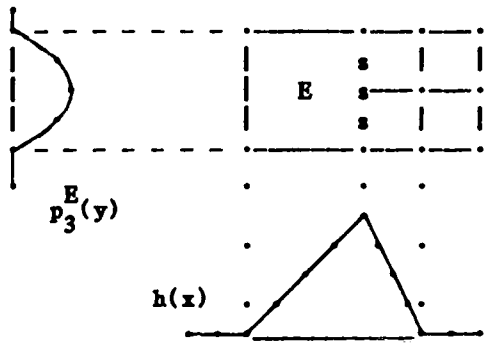


Figure 3-1: A side basis function

The <u>element</u> basis functions for an unrefined element E are the functions

$$p_i^E(x)\ p_m^E(y) : i,m \geq 2,\ i+m-1 \leq k_E\ . \tag{3.4}$$

Element basis functions for E are nonzero only in E, and only occur when $k_E \geq 5$.

Since the mesh is 1-irregular, <u>each</u> basis function B_I for S is of the form $p_i^{Ex}(x)\ p_m^{Ey}(y)$ in each unrefined element E, where Ex is either E or f(E), and Ey is either E or f(E). Moreover, (3.3) and (3.4) imply that in each element E, any monomial $x^i y^m$ with $i+m-1 \leq k_E$ is a linear combination of restrictions to E of basis functions, so (3.1) holds.

Figure 3-2 depicts the values of the basis functions in an element in a sample local-order finite element. Note that many of the basis functions are of order greater than $k_E = 3$.

3.3 The Hermite case

The construction in the last section extends immediately to spaces of globally $C^{\alpha-1}$ functions for any integer $\alpha \geq 1$. The hierarchic polynomials satisfy Hermite interpolation conditions on their first α derivatives at 0 and 1, so $k_E \geq 2\alpha$. Vertex, side, and element basis functions are defined as in the C^0 case. There are

α^2 vertex basis functions for each regular vertex,

$(k_E-4\alpha)(k_E-4\alpha+1)/2$ element basis functions for element E (none if $k_E \leq 4\alpha$),

and

$\sum_{i=1}^{\alpha} (k_s-2\alpha+1-i)$ side basis functions for side s,

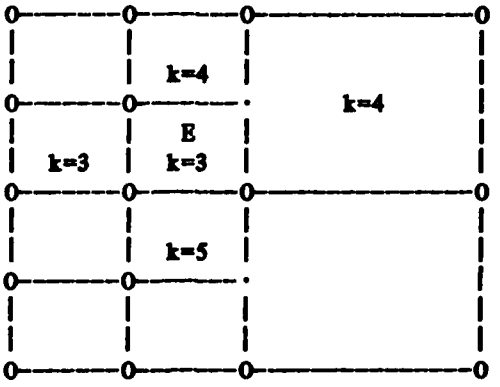
and condition (3.1) holds as before.

3.4 A mesh data structure for local order refinement

In this section, we extend the mesh data structure from Section 2.5 to handle local-order finite element spaces.

For each unrefined element E, we store $-k_E$ in IRCT, in the location in the node for f(E) which will contain a pointer to the node for E if E is refined.

mesh :



basis functions in element E :

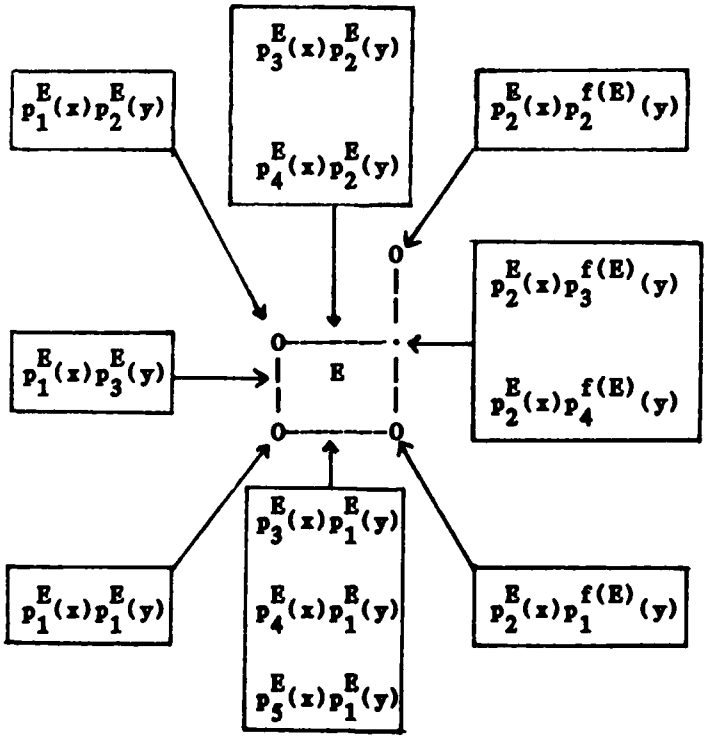


Figure 3-2: Local-order basis functions

Let the function MBV map the basis functions for S to the regular

vertices in the mesh, such that

1. MBV maps each vertex basis function to its vertex.

2. MBV maps each side basis function for side s to the last-created regular vertex at an end of s.

3. MBV maps each element basis function for element E to the center vertex of f(E).

A sample mesh indicating MBV is shown in Figure 3-3.

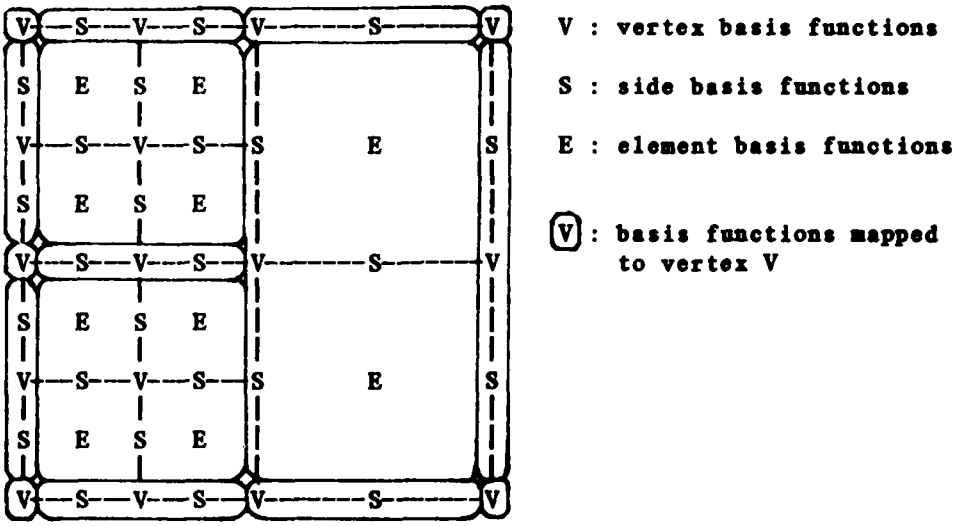


Figure 3-3: Basis functions mapped to vertices

We order the vertices and basis functions for S so that the basis functions mapped by MBV to vertex V_{IV} are

$$B_{MVB(IV)}, B_{MVB(IV)+1}, \cdots, B_{MVB(IV+1)-1},$$

and we associate a number IBFT(IBF) with each basis function B_{IBF}, encoding the spatial relation to the vertex MBV(B_{IBF}) (e.g., left, lower right, etc.).

With the data structure from Section 2.5, and two additional arrays containing MVB and IBFT, we can easily determine the basis functions which are nonzero in a given unrefined element.

As in Chapter 2, this mesh data structure makes it relatively easy to construct the finite element system for S, to evaluate the approximate solution, and to refine the mesh.

CHAPTER 4

A posteriori error estimators

4.1 Introduction

The traditional error estimates for the finite element method for elliptic problems are a priori (e.g., Ciarlet [17]), predicting the expected rate of convergence of the error to zero, but not saying much about the actual error for a given finite element space.

In recent years, locally-computable a posteriori error estimators have been developed, mainly by Babuška and Rheinboldt, which estimate the actual error for a given space. The landmark paper in this area is by Babuška and Rheinboldt [8], giving a method of constructing a locally-computable error estimator which is within a multiplicative constant of the norm of the actual error. Further results for one-dimensional problems are given by Babuška and Rheinboldt [6, 7, 5] and Reinhardt [24]. Further results for two-dimensional problems are given by Babuška and Miller [4]: under suitable assumptions, they prove that the error estimator used in their code (with piecewise bilinear basis functions) converges to the norm of the actual error as the mesh size goes to zero.

In Section 4.2, we present some terminology for discussing a posteriori error estimators for two-dimensional Neumann problems. In Section 4.3, we briefly describe the error estimator presented in [8]. In Section 4.4, we present a new error estimator which has two main advantages over the estimator in [8]:

1. Under reasonable assumptions, the new error estimator is asymptotically an upper bound on the norm of the true error.

2. The new error estimator can be computed locally in each element. (The one in [8] must be computed over more complicated local regions.)

In Section 4.5, we prove the upper-bound property of the new error estimator, and, under suitable assumptions, we show that the new error estimator is no larger than a multiplicative constant times the norm of the actual error. In Section 4.6, we present several cheaper alternative error estimators. In Section 4.7, we extend the estimators to handle more general boundary conditions and differential operators. In section 4.8, we summarize the results in this chapter.

In Chapter 7, we present numerical evidence that, in some cases, our error estimators converge to the norm of the actual error as the mesh size goes to zero. We have not been able to prove this convergence. Currently, we can only prove convergence for our error estimators for one-dimensional problems [15].

4.2 Terminology

Consider the model Neumann problem

$$Lu := - D_x(a(x,y)D_x u) - D_y(a(x,y)D_y u) + b(x,y)u = f(x,y) \qquad (4.1)$$

$$\text{in the interior of } \Omega,$$

$$\partial u/\partial n = 0 \quad \text{on } \partial\Omega,$$

where D_x , D_y , and $\partial/\partial n$ denote derivatives in x, y, and the outward normal to Ω respectively, a(x,y) and b(x,y) are in $L^\infty(\Omega)$, f(x,y) is in $L^2(\Omega)$, $0 < \underline{a} \le a(x,y) \le \bar{a}$, and $0 < \underline{b} \le b(x,y) \le \bar{b}$. Variants of (4.1) are discussed in Section 4.7.

Let ω be an open subset of Ω with piecewise smooth boundary $\partial\omega$. For a non-negative integer k, and a function v in $C^\infty(\omega)$, we define the norm

$$||v||^2_{k,\omega} := \sum_{i=0}^{k} \sum_{j=0}^{i} \iint_\omega (D_x^j D_y^{i-j} v(x,y))^2 \, dx \, dy.$$

Let $H^k(\omega)$ denote the completion of $C^\infty(\omega)$ with respect to the norm $||\cdot||_{k,\omega}$. $H^k(\omega)$ is equivalent to

$$\{v \text{ in } L^2(\omega) : ||v||_{k,\omega} < \infty\},$$

if derivatives are defined in the distributional sense (e.g., Ciarlet [17], page 114). Let $H_0^k(\omega)$ denote the completion of the functions in $C^\infty(\omega)$ with compact support in ω with respect to the norm $||\cdot||_{k,\omega}$. When k is omitted, k = 0 is assumed. For v,w in $H^0(\omega)$

$(= L^2(\omega))$, let

$$(v,w)_\omega := \iint_\omega v(x,y)\, w(x,y)\, dx\, dy.$$

For v,w in $H^1(\omega)$, let

$$a_\omega(v,w) := (aD_x v, D_x w)_\omega + (aD_y v, D_y w)_\omega + (bv,w)_\omega .$$

If ω is the interior of Ω, $a_\omega(u,v)$ is obtained by multiplying (4.1) by a function v, and integrating over ω using Green's theorem. Let the <u>energy norm</u> $|||v|||_\omega^2 := a_\omega(v,v)$. By the assumptions on $a(x,y)$ and $b(x,y)$, there exists $C > 0$* independent of v, such that

$$C^{-1} \|v\|_{1,\omega} \le |||v|||_\omega \le C \|v\|_{1,\omega} \tag{4.2}$$

for all v in $H^1(\omega)$.

On the boundary $\partial\omega$, for v and w in $L^2(\partial\omega)$, let

$$\langle v,w\rangle_{\partial\omega} := \int_{\partial\omega} v(\sigma)\, w(\sigma)\, d\sigma,$$

where $d\sigma$ is arc length. We define the norm

---

*C denotes a generic constant, not necessarily the same in each instance.

$$|v|^2_{\partial\omega} := \langle v,v\rangle_{\partial\omega} .$$

Let $\partial/\partial n$ denote differentiation in the direction of the outward normal to ω, with the limit in the differentiation taken over points inside ω.

For brevity, we sometimes replace ω by its closure in names for spaces, norms, and inner products: for example, $||v||_{k,E}$ stands for $||v||_{k,\text{interior}(E)}$. When the subscript ω is omitted in norms and inner products, ω = interior(Ω) is assumed.

Given a 1-irregular mesh M with unrefined elements {E}, we choose a finite element subspace S (= S(M)) of $H^1(\Omega)$ as in Chapters 2 and 3. For simplicity, we assume S is based on C^0 hierarchic polynomials, as in Section 3.2. We also choose a larger finite element space $\bar{S}$ (= $\bar{S}$(M)) containing S, constructed in the same manner as S, such that in each unrefined element E,

smooth functions are locally approximated to order $\bar{k}_E > k_E$.

Since $\bar{S}$ contains S, and the basis functions $\{B_I\}$ are hierarchic, the basis

$$\{B_I : B_I \text{ in } \bar{S}\}$$

for $\bar{S}$ contains the basis

$$\{B_I : B_I \text{ in } S\}$$

for S. For each unrefined element E, we define $\bar{S}_E$ to be the space with basis functions

$$\{B_I|_E : E \text{ in } \mathrm{supp}(B_I),\ B_I \text{ in } \bar{S}\},$$

where $\{B_I|_E : E \text{ in } \mathrm{supp}(B_I)\}$ are linearly independent because of (2.2), (3.3), and (3.4), and we define S_E to be the space with basis functions

$$\{B_I|_E : E \text{ in } \mathrm{supp}(B_I),\ B_I \text{ in } S\}.$$

We define $\bar{S}_E - S_E$ to be the span of the functions

$$\{B_I|_E : E \text{ in } \mathrm{supp}(B_I),\ B_I \text{ in } \bar{S},\ B_I \text{ not in } S\}.$$

For any function

$$v = \sum_{B_I \in \bar{S}} v_I\, B_I|_E$$

in $\bar{S}_E$, we define

$$I_E(v) := \sum_{B_I \in S} v_I\, B_I|_E \tag{4.3}$$

in S_E. Then $v - I_E(v)$ is in $\bar{S}_E - S_E$, and for any function

$$v = \sum_{B_I \in \bar{S}} v_I\, B_I$$

in $\bar{S}$, the function

$$I(v) := \sum_{B_I \in S} v_I\, B_I$$

is in S, and is equal to $I_E(v)$ in E.

Let N(E) denote the set of unrefined elements which intersect with E at more than one point. For example, in Figure 3-2, N(E) is the set of labelled elements.

For convenience in discussing functions in the product space

$$H_M(\Omega) := \{v: v \text{ in } H^1(E) \text{ for all } E\},$$

let

$$a(v,w) := \sum_E a_E(v,w) \quad \text{for } v \text{ and } w \text{ in } H_M(\Omega),$$

and

$$|||v|||^2 := \sum_E |||v|||_E^2 \quad \text{for } v \text{ in } H_M(\Omega).$$

The weak form of (4.1),

$$a(u,v) = (f,v) \quad \text{for all } v \text{ in } H^1(\Omega),$$

has a unique <u>solution</u> u in $H^1(\Omega)$ (e.g., Ciarlet [17], page 19). We impose the weak form of (4.1) in S to get a <u>finite element approximation</u> U in S, where

$$a(U,v) = (f,v) \quad \text{for all } v \text{ in } S.$$

The <u>error</u> for S is $e := u-U$. Similarly, $\bar{U}$ in $\bar{S}$ is required to satisfy

$$a(\bar{U},v) = (f,v) \quad \text{for all } v \text{ in } \bar{S},$$

and the corresponding error for $\bar{S}$ is $\bar{e} := u-\bar{U}$.

In this chapter, we consider ways of estimating the norm of the error $|||e|||$. Let

$$\varepsilon := \bar{U}-U.$$

We assume the <u>saturation condition</u> that there exists $0 \leq \bar{\gamma} < 1$ such that

$$|||\bar{e}||| \leq \bar{\gamma}\, |||e|||. \tag{4.4}$$

If u is sufficiently smooth, then $\bar{\gamma}$ goes to zero as h goes to zero. Thus, since

$$|||\varepsilon||| \leq |||e||| \leq (1-\bar{\gamma})^{-1}\, |||\varepsilon|||,$$

we concentrate our attention on estimating $|||\varepsilon|||$.

4.3 A Dirichlet a posteriori error estimator

Traditionally, the error estimates for the finite element method have been a priori, predicting the expected rate of convergence of $|||e|||$ to zero (usually as $h := \max\{h_E\}$ goes to zero), but not saying much about the actual error for a given S.

In [8], Babuška and Rheinboldt present a posteriori error estimators constructed basically as follows. Let $\{B_J\}$ be a subset of the basis functions for S which form a partition of unity on Ω. For each J, let $\Omega_J := \text{supp}(B_J)$, and let e_J be the solution in $H_0^1(\Omega_J)$ of the local Dirichlet problem

$$a_{\Omega_J}(e_J,v) = (f,v)_{\Omega_J} - a_{\Omega_J}(U,v) \quad \text{for all } v \text{ in } H_0^1(\Omega_J).$$

Under suitable assumptions, Babuška and Rheinboldt show that there exist $C_1 > 0$ and $C_2 > 0$ independent of S such that

$$C_1 \, |||e|||^2 \leq \sum_J |||e_J|||^2_{\Omega_J} \leq C_2 \, |||e|||^2.$$

In practice, since $H_0^1(\Omega_J)$ is infinite-dimensional, e_J is not computable. The most straightforward computable approximation of e_J is ε_J in

$$\bar{S}_J := \{p(x,y) : p \text{ in } \bar{S}, \text{ supp}(p) \text{ contained in } \Omega_J\}$$

defined by the equations

$$a_{\Omega_J}(\varepsilon_J,v) = (f,v)_{\Omega_J} - a_{\Omega_J}(U,v) \quad \text{for all } v \text{ in } \bar{S}_J \, . \tag{4.5}$$

Then

$$C_1 \, |||\varepsilon|||^2 \leq \sum_J |||\varepsilon_J|||^2_{\Omega_J} \leq C_2 \, |||\varepsilon|||^2 \, . \tag{4.6}$$

The quantity $\left(\sum_J |||\varepsilon_J|||^2_{\Omega_J} \right)^{1/2}$ is called an <u>error estimator</u>.

Each ε_J in (4.5) is a projection of ε in the energy norm onto a set of functions, so the right hand inequality in (4.6) arises naturally, while the left hand inequality is harder to prove.

4.4 A Neumann a posteriori error estimator

The motivation for a Neumann a posteriori error estimator is that it is more important to have an upper bound on $|||\varepsilon|||$ than a lower bound. We construct an error simulator $\tilde{\varepsilon}$, such that ε is the projection of $\tilde{\varepsilon}$ onto $\bar{S}$ with respect to the energy norm. Then $|||\varepsilon||| \leq |||\tilde{\varepsilon}|||$ arises naturally. Following Babuška and Rheinboldt [10], we call $|||\tilde{\varepsilon}|||$ an error estimator, and for each element E, we call $|||\tilde{\varepsilon}|||_E$ an error indicator.

To motivate our construction of $\tilde{\varepsilon}$, suppose u is in $H^2(E)$. By Green's theorem on element E,

$$a_E(e,v) = F_E'v) := (f,v)_E + \langle a\partial u/\partial n, v\rangle_{\partial E} - a_E(U,v) \qquad (4.7)$$

for all v in $H^1(E)$. For a given v, we can compute every quantity in (4.7) except $a\partial u/\partial n$ on the boundary of E. Suppose E shares side s with element E'. On s, a possible approximation to $a\partial u/\partial n|_E$ is the average value of $a\partial U/\partial n_E$ on E and E', where n_E is the outward normal from E. Since

$$n_E|_s = - n_{E'}|_s \ ,$$

we define the approximation

$$[a\partial U/\partial n]|_E := \begin{cases} 1/2 \ \{ \ a\partial U/\partial n|_E - a\partial U/\partial n|_{E'} \ \} \ , & s \text{ not in } \partial\Omega \\ 0 \ , & s \text{ in } \partial\Omega \ . \end{cases} \qquad (4.8)$$

Note

$$[a\partial U/\partial n]|_E = -[a\partial U/\partial n]|_{E'} .$$

The resulting approximation to $F_E(v)$ is

$$\tilde{F}_E(v) := (f,v)_E + \langle [a\partial U/\partial n], v\rangle_{\partial E} - a_E(U,v). \tag{4.9}$$

Since $a_E(\cdot,\cdot)$ is elliptic, we can define an error simulator for e in E by replacing F_E with $\tilde{F}_E$ on the right hand side of (4.7). However, as h_E approaches zero, the finite element matrices for the error simulator approach an indefinite matrix corresponding to the case when $b(x,y) = 0$.

Thus, we formulate error simulator equations which are automatically consistent when $b(x,y) = 0$ and $a_E(\cdot,\cdot)$ is indefinite. We approximate ε by the function $\tilde{\varepsilon}_E$ in $\bar{S}_E$ defined by the equations

$$a_E(\tilde{\varepsilon}_E, v) = \tilde{F}_E(v - I_E(v)) \quad \text{for all } v \text{ in } \bar{S}_E , \tag{4.10}$$

where $I_E(v)$ is defined by (4.3). Since $a_E(\cdot,\cdot)$ is elliptic, $\tilde{\varepsilon}_E$ exists uniquely. Since v need not vanish on ∂E, (4.10) is a Neumann problem. Moreover, since $1|_E$ is in S_E ,

$$I_E(1) = 1,$$

so the consistency condition $a_E(\tilde{\varepsilon}_E, 1) = 0$ is automatically satisfied.

Equations (4.10) are equivalent to choosing $\tilde{\varepsilon}_E$ in $\bar{S}_E$ such that

$$a_E(\tilde{\varepsilon}_E, v) = \begin{cases} \tilde{F}_E(v) & \text{for } v \text{ in } \bar{S}_E - S_E \\ 0 & \text{for } v \text{ in } S_E \end{cases} .$$

The corresponding matrix equation is $\tilde{A}\,\tilde{v} = f$ for the coefficient vector $\tilde{v}$ of $\tilde{\varepsilon}_E$, where

$$\tilde{A} := \begin{bmatrix} \tilde{A}_1 & \tilde{A}_3 \\ \tilde{A}_3^T & \tilde{A}_2 \end{bmatrix} , \quad \tilde{v} := \begin{bmatrix} \tilde{v}_1 \\ \tilde{v}_2 \end{bmatrix} , \quad f := \begin{bmatrix} 0 \\ f_2 \end{bmatrix} ,$$

the coefficients of the basis functions for S_E are in $\tilde{v}_1$, and the coefficients of the basis functions for $\bar{S}_E - S_E$ are in $\tilde{v}_2$.

We define the value of the <u>error simulator</u> $\tilde{\varepsilon}$ at (x,y) in Ω to be

$$\tilde{\varepsilon}(x,y) := \tilde{\varepsilon}_E(x,y),$$

where E is an unrefined element containing (x,y). When there is no danger of confusion, we refer to $\tilde{\varepsilon}_E$ as $\tilde{\varepsilon}$. Note that $\tilde{\varepsilon}$ is in $H_M(\Omega)$, but not necessarily in $H^1(\Omega)$.

4.5 Properties of the Neumann error estimator

The most important property of $\tilde{\varepsilon}$ is that

<u>Theorem 4.1</u>: $|||\varepsilon||| \leq |||\tilde{\varepsilon}|||$.

<u>Proof</u>: Suppose v is in $\bar{S}$. Since $[a\partial U/\partial n]|_E = -[a\partial U/\partial n]|_E$, , and $[a\partial U/\partial n] = 0$ on $\partial\Omega$,

$$\sum_E \langle [a\partial U/\partial n], v\rangle_{\partial E} = 0.$$

Thus,

$$\sum_E \tilde{F}_E(v) = \sum_E \{(f,v)_E - a_E(U,v)\} = (f,v) - a(U,v).$$

Since I(v) (page 38) is in S,

$$\sum_E \tilde{F}_E(I_E(v)) = \sum_E \tilde{F}_E(I(v)) = (f,I(v)) - a(U,I(v)) = 0. \qquad (4.11)$$

Thus,

$$a(\varepsilon,v) = a(e,v) = (f,v) - a(U,v) = \sum_E \tilde{F}_E(v-I_E(v)) = a(\tilde{\varepsilon},v), \quad (4.12)$$

and, in particular,

$$|||\varepsilon|||^2 = a(\varepsilon,\varepsilon) = a(\tilde{\varepsilon},\varepsilon) \leq |||\tilde{\varepsilon}||| \; |||\varepsilon|||.$$

By (4.4), we have

Corollary 4.2: $(1-\bar{\gamma})\; |||e||| \leq |||\tilde{\varepsilon}|||.$

This is a global inequality. We cannot generally expect $|||e|||_E \leq C\; |||\tilde{\varepsilon}|||_E$: suppose u is smooth, S contains bilinears, $h_E \cong \underline{h}$ in region R in Ω, and $h_E \cong \bar{h} \gg \underline{h}$ away from R. In region R, U is a finite element approximation to the solution to

$$L(u_R) = f(x,y) \quad \text{in the interior of R,} \qquad (4.13)$$
$$u_R = U \qquad \text{on } \partial R.$$

Let E be an element in the interior of R. Since the error simulator equations in E are <u>local</u>, they are the same for (4.13) as for (4.1). Since (Theorem 4.6) $|||\tilde{\varepsilon}|||_E$ is bounded by local quantities behaving like $|||U-u_R|||_E$, and the boundary error in (4.13) pollutes the solution throughout R, we expect

$$|||e|||_E \gg |||\tilde{\varepsilon}|||_E .$$

(In practice, this may be acceptable, since the error is bigger away from R due to the coarse mesh, and the small size of the error simulator is in accord with the fact that no more mesh refinement should occur.)

By the triangle inequality and (4.12),

$$|||\tilde{\varepsilon}||| - |||\varepsilon||| \leq |||\tilde{\varepsilon}-\varepsilon||| = \inf\{|||\tilde{\varepsilon}-f|||: f \text{ in } \bar{S}\},$$

so we can get an a posteriori upper bound for $|||\tilde{\varepsilon}|||-|||\varepsilon|||$ by measuring the difference in the energy norm between $\tilde{\varepsilon}$ and any function in $\bar{S}$. For instance, by averaging $\tilde{\varepsilon}$, we can construct an approximation $\tilde{\varepsilon}_{ave}$ to $\tilde{\varepsilon}$ in $\bar{S}$: $\tilde{\varepsilon}-\tilde{\varepsilon}_{ave}$ approximates the size of the jumps in $\tilde{\varepsilon}$ across element boundaries. If

$$|||\tilde{\varepsilon}-\tilde{\varepsilon}_{ave}|| \leq \gamma\ |||\tilde{\varepsilon}|||$$

with $\gamma < 1$, then

$$|||\tilde{\varepsilon}||| \leq (1-\gamma)^{-1}\ |||\varepsilon|||.$$

Of course, the jumps in $\tilde{\varepsilon}$ can be made as small as desired by adding

penalty terms to equations (4.10), but the resulting equations are non-local.

We now prove a <u>local</u> a priori upper bound for $|||\tilde{\varepsilon}_E|||_E$ in terms of quantities depending on e. We assume there exists a constant kmax such that

$$\bar{k}_E \leq \text{kmax independent of } \bar{S}. \tag{4.14}$$

We start with a few lemmas.

<u>Lemma 4.3</u>: $|||I_E(\tilde{\varepsilon}_E)|||_E \leq C\,|||\tilde{\varepsilon}_E|||_E$.

<u>Proof</u>: Suppose v is in $\bar{S}_E$. Since the dimension of $\bar{S}_E$ is finite and I_E is linear,

$$(bI_E(v),I_E(v))_E \leq C_1\,(bv,v)_E ,$$

where C_1 is independent of v and h_E . (Otherwise, there would exist w in $\bar{S}_E$ such that $(bw,w)_E = 0$ and $(bI_E(w),I_E(w))_E \neq 0$, which leads to a contradiction.) Similarly, since the dimension of $\bar{S}_E$ is finite, I_E is linear, and $I_E(1) = 1$,

$$(aD_xI_E(v),D_xI_E(v))_E + (aD_yI_E(v),D_yI_E(v))_E \leq$$

$$C_2\,(\,(aD_xv,D_xv)_E + (aD_yv,D_yv)_E\,),$$

where C_2 is independent of v and h_E . Thus,

$$|||I_E(v)|||_E^2 \leq \min(C_1,C_2)\ |||v|||_E^2 .$$

Let $v = \tilde{\varepsilon}_E$.

Lemma 4.4: $|\tilde{\varepsilon}_E - I_E(\tilde{\varepsilon}_E)|_{\partial E} \leq C\ h_E^{1/2} |||\tilde{\varepsilon}_E|||_E$.

Proof: The trace inequality

$$|\tilde{\varepsilon}_E - I_E(\tilde{\varepsilon}_E)|_{\partial E}^2 \leq C\ (h_E^{-1}\ ||\tilde{\varepsilon}_E - I_E(\tilde{\varepsilon}_E)||_E^2 + h_E\ ||\tilde{\varepsilon}_E - I_E(\tilde{\varepsilon}_E)||_{1,E}^2) \quad (4.15)$$

follows directly from Theorem 3.10 in Agmon [1].

Suppose v is in $\bar{S}_E$. For $0 \leq x,y \leq 1$, let

$$v_\Omega(x,y) := v(x_E + h_E\, x,\ y_E + h_E\, y),$$

$$I_\Omega(x,y) := I_E(v)(x_E + h_E\, x,\ y_E + h_E\, y).$$

Since the dimension of $\bar{S}_E$ is finite, I_E is linear, and $I_E(1) = 1$,

$$||v_\Omega - I_\Omega||_\Omega^2 \leq C\ (||D_x v_\Omega||_\Omega^2 + ||D_y v_\Omega||_\Omega^2),$$

and thus

$$||v - I_E(v)||_E^2 \leq C\ h_E^2\ ||v||_{1,E}^2 ,$$

where C is independent of v and h_E . The result follows by letting $v = \tilde{\varepsilon}_E$ in (4.15) and using (4.2) and Lemma 4.3.

Lemma 4.5: If u is in $H^2(E)$,

$$|a\partial e/\partial n|_{\partial E} \leq C\, h_E^{-1/2} \{ \|e\|_{1,E}^2 + h_E^2 \|e\|_{2,E}^2 \}^{1/2}.$$

Proof: Since u is in $H^2(E)$, e is in $H^2(E)$. The result follows directly from Theorem 3.10 in Agmon [1].

The following theorem shows that the error is large in local regions where the error indicators are large (and u is sufficiently smooth).

Theorem 4.6: If u is in $H^2(E')$ for all E' in N(E), then

$$|||\tilde{\varepsilon}_E|||_E \leq C \sum_{E' \in N(E)} \{ \|e\|_{1,E'}^2 + h_{E'}^2 \|e\|_{2,E'}^2 \}^{1/2}.$$

Proof: For any v in $\bar{S}_E$,

$$a_E(\tilde{\varepsilon}_E, v) = (f, v - I_E(v))_E + \langle [a\partial U/\partial n], v - I_E(v) \rangle_{\partial E} - a_E(U, v - I_E(v))$$

$$= F_E(v - I_E(v)) - \langle a\partial u/\partial n, v - I_E(v) \rangle_{\partial E} + \langle [a\partial U/\partial n], v - I_E(v) \rangle_{\partial E}$$

$$= a_E(e, v - I_E(v)) - \langle [a\partial e/\partial n], v - I_E(v) \rangle_{\partial E}.$$

Let $v = \tilde{\varepsilon}_E$. By the Cauchy-Schwarz inequality and the definition of $[\cdot]$,

$$|||\tilde{\varepsilon}_E|||_E^2 \leq |||e|||_E \, |||\tilde{\varepsilon}_E - I_E(\tilde{\varepsilon}_E)|||_E +$$

$$\{ \sum_{E' \in N(E)} |a\partial e/\partial n|_{\partial E'} \} \, |\tilde{\varepsilon}_E - I_E(\tilde{\varepsilon}_E)|_{\partial E}.$$

By Lemma 4.3 and Lemma 4.4,

$$|||\tilde{\varepsilon}_E|||_E \leq C\,|||e|||_E + C\,h_E^{1/2} \sum_{E' \in N(E)} |a\partial e/\partial n|_{\partial E'} . \tag{4.16}$$

By Lemma 4.5, noting that (since the mesh is 1-irregular) $h_E^{1/2}\, h_{E'}^{-1/2} \leq C$ for all E' in $N(E)$,

$$|||\tilde{\varepsilon}_E|||_E \leq C\,|||e|||_E +$$

$$C \sum_{E' \in N(E)} \{\, \|e\|^2_{1,E'} + h^2_{E'}\, \|e\|^2_{2,E'} \,\}^{1/2}.$$

The result follows by noting that E is in N(E) and using (4.2).

If u is in $H^2(E)$, we expect the right hand side of Theorem 4.6 to converge to zero at the same rate as $|||e|||_E$. If the coefficients a(x,y) and b(x,y) are sufficiently smooth, u is in $H^2(\Omega)$ (e.g., Ciarlet [17], page 138).

We now prove a <u>global</u> a priori upper bound for $|||\tilde{\varepsilon}|||$. We assume that u is in $H^2(E)$ for all E, and we assume the <u>saturation condition</u> that there exists $0 \leq \bar{\bar{\gamma}} < \infty$ such that

$$\{ \sum_E h_E\, |a\partial\bar{e}/\partial n|^2_{\partial E} \}^{1/2} \leq \bar{\bar{\gamma}}\, |||e|||. \tag{4.17}$$

If u is sufficiently smooth, then $\bar{\bar{\gamma}}$ goes to zero if h goes to zero.

<u>Theorem 4.7</u>: There exists $\tilde{C}$ independent of S and $\bar{S}$ such that

$$|||\tilde{\varepsilon}||| \leq \tilde{C}\, |||e|||.$$

Proof: It follows from Theorem 3.10 in Agmon [1] and (4.14) that

$$|a\partial\varepsilon/\partial n|^2_{\partial E} \le C\, h_E^{-1} \{ \|\varepsilon\|^2_{1,E} + h_E^2 \|\varepsilon\|^2_{2,E} \}$$

$$\le C\, h_E^{-1} \|\varepsilon\|^2_{1,E} .$$

Then, using $a\partial e/\partial n = a\partial\bar{e}/\partial n + a\partial\varepsilon/\partial n$ in Equation (4.16),

$$|||\tilde{\varepsilon}_E|||_E \le C\, |||e|||_E + C\, h_E^{1/2} \sum_{E' \in N(E)} |a\partial\bar{e}/\partial n|_{\partial E'} +$$

$$C \sum_{E' \in N(E)} \|\varepsilon\|_{1,E'} .$$

Squaring, summing over E, using the triangle inequality, and noting that the maximum number of $N(\cdot)$'s in which any element appears is bounded,

$$|||\tilde{\varepsilon}||| \le C\, |||e||| + C\, \|\varepsilon\|_1 + C \{ \sum_E h_E |a\partial\bar{e}/\partial n|^2_{\partial E} \}^{1/2}.$$

The Theorem follows using (4.2), noting that $|||\varepsilon||| \le |||e|||$, and using (4.17).

4.6 Alternative error estimators

In practice, there are several alternative error estimators that are less costly to compute than $|||\tilde{\varepsilon}|||$.

4.6.1 Ignoring irregular corners

We can compute an error simulator in element E by replacing $I_E(v)$ with an analogous function $I^*_E(v)$ which interpolates v locally in E. The resulting error simulator, in effect, ignores the irregularity of the corners of E, and its system is slightly cheaper to assemble than the

system $\tilde{A}\,\tilde{v} = f$. In practice, the resulting error estimator appears to be at least as accurate as $|||\tilde{s}|||$. However, Equation (4.11) in the proof of Theorem 4.1 break down when v is a side basis function for a side s containing an irregular vertex (Figure 4-1): $I^*(v)$, the analogue to I(v), is not continuous across s. Thus, we do not consider this alternative error estimator further.

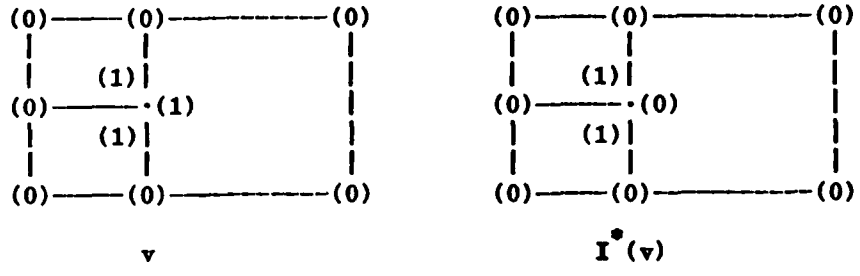


Figure 4-1: Ignoring the irregular corners

4.6.2 Using the matrix for a local Poisson problem

We can form and solve error simulator equations using a matrix corresponding to an associated local Poisson problem. For v and w in $H^1(E)$, let

$$\underline{a}_E(v,w) := \underline{a}(E)\ \{\ (D_x v, D_x w)_E + (D_y v, D_y w)_E\ \},$$

$$|||v|||_E^2 := \underline{a}_E(v,v),$$

where $\underline{a}(E) := \inf\{a(x,y) \text{ in } E\}$.

Instead of choosing $\tilde{e}_E$ in $\bar{S}_E$ by (4.10), we can choose $\underline{e}_E$ in $\bar{S}_E$ such that

$$a_E(\underline{\varepsilon}_E, v) = \begin{cases} \tilde{F}_E(v) & \text{for all } v \text{ in } \bar{S}_E - S_E \\ 0 & \text{for all } v \text{ in } S_E \ , \end{cases} \tag{4.18}$$

and

$$(\underline{\varepsilon}_E, 1)_E = 0. \tag{4.19}$$

(We impose (4.19) because $a_E(1,1) = 0$.) We construct the error simulator $\underline{\varepsilon}$ and the error indicators $\{|||\underline{\varepsilon}|||_E\}$ from $\underline{\varepsilon}_E$ in the same way we constructed $\tilde{\varepsilon}$ and $\{|||\tilde{\varepsilon}|||_E\}$ from $\tilde{\varepsilon}_E$.

Equations (4.18) correspond to the matrix equation $\underline{A}\,\underline{v} = f$ for the coefficient vector $\underline{v}$ of $\underline{\varepsilon}_E(x,y)$, where

$$\underline{A} := \begin{bmatrix} A_1 & A_3 \\ A_3^T & A_2 \end{bmatrix} , \quad \underline{v} := \begin{bmatrix} \underline{v}_1 \\ \underline{v}_2 \end{bmatrix} , \quad f := \begin{bmatrix} 0 \\ f_2 \end{bmatrix} ,$$

the coefficients of the basis functions for S_E are in $\underline{v}_1$, and the coefficients of the basis functions for $\bar{S}_E - S_E$ are in $\underline{v}_2$. Since $|||\underline{\varepsilon}|||_E = |||\underline{\varepsilon}+C|||_E$ for any constant C, we can set

$$A_{1,J} = A_{J,1} = \delta_{J,1} \ ,$$

reordering the unknowns if necessary, so that $\underline{A}$ is positive definite. For fixed k_E , $\underline{A}$ depends only on the regularity of the corners of E, and not on h_E .

Theorem 4.8: If $a(x,y)$ is Lipschitz-continuous, then

$$|||\tilde{\varepsilon}|||_E \leq |\underline{|}\underline{\varepsilon}\underline{|}|_E \leq (1+O(h_E))\ |||\tilde{\varepsilon}|||_E$$

as h_E goes to zero.

Proof: By (4.10) and (4.18), since $\underline{a}_E(v,v) \leq a_E(v,v)$ for all v in $\bar{S}_E$,

$$|||\tilde{\varepsilon}|||_E^2 = \underline{a}_E(\underline{\varepsilon},\tilde{\varepsilon}) \leq |\underline{|}\underline{\varepsilon}\underline{|}|_E\ |\underline{|}\tilde{\varepsilon}\underline{|}|_E \leq |\underline{|}\underline{\varepsilon}\underline{|}|_E\ |||\tilde{\varepsilon}|||_E .$$

Similarly,

$$|\underline{|}\underline{\varepsilon}\underline{|}|_E^2 = \underline{a}_E(\underline{\varepsilon},\underline{\varepsilon}) = a_E(\tilde{\varepsilon},\underline{\varepsilon}) \leq |||\tilde{\varepsilon}|||_E\ |||\underline{\varepsilon}|||_E ,$$

and if $a(x,y)$ is Lipschitz-continuous in E, so that $a(x,y) \leq \underline{a}(E) + O(h_E)$ in E, then by (4.19),

$$\|\underline{\varepsilon}\|_E \leq C\ h_E\ \|\underline{\varepsilon}\|_{1,E} ,$$

and

$$|||\underline{\varepsilon}|||_E \leq (1+O(h_E))\ |\underline{|}\underline{\varepsilon}\underline{|}|_E$$

as h_E goes to zero.

Since mesh refinement occurs in a very regular fashion, we can precompute the factorization of $\underline{A}$, and only incur the cost of a backsolve plus computation of the right hand side for each element. If the hierarchic polynomials $\{p_i\}$ are chosen by (3.2), then since

$$\int_0^1 p_i' p_j' = 0 , \quad i \neq j , \quad \max(i,j) > 2,$$

$\underline{A}$ is sparse. In practice, we replace the quantity $\underline{a}(E)$ with the more convenient quantity $a(x_E+h_E/2, y_E+h_E/2)$. Then

$$|||\tilde{\varepsilon}|||_E \leq (1+O(h_E))\ |\underline{|}\underline{|}\underline{\varepsilon}|\underline{|}\underline{|}_E \leq (1+O(h_E))\ |||\tilde{\varepsilon}|||_E \ .$$

4.6.3 Using a smaller matrix

We can also compute $\bar{\varepsilon}_E$ in $\bar{S}_E - S_E$ by the equations

$$\underline{a}_E(\bar{\varepsilon}_E, v) = \tilde{F}_E(v) \quad \text{for all } v \text{ in } \bar{S}_E - S_E \ , \tag{4.20}$$

constructing the error simulator $\bar{\varepsilon}$ and the error indicators $\{|\underline{|}|\bar{\varepsilon}|\underline{|}|_E\}$ from $\bar{\varepsilon}_E$ in the same way we constructed $\tilde{\varepsilon}$ and $\{|||\tilde{\varepsilon}|||_E\}$ from $\tilde{\varepsilon}_E$. Equations (4.20) correspond to the matrix equation $A_2\ \bar{v}_2 = f_2$, or $\bar{A}\ \bar{v} = f$, for the coefficient vector $\bar{v}$ of $\bar{\varepsilon}_E$, where

$$\bar{A} := \begin{bmatrix} A_1 & 0 \\ 0 & A_2 \end{bmatrix} \ , \quad \bar{v} := \begin{bmatrix} 0 \\ \bar{v}_2 \end{bmatrix} \ , \quad f := \begin{bmatrix} 0 \\ f_2 \end{bmatrix} \ .$$

We now show

Theorem 4.9: There exists a constant $\bar{C}$ independent of h_E such that

$$|\underline{|}|\bar{\varepsilon}|\underline{|}|_E \leq |\underline{|}|\underline{\varepsilon}|\underline{|}|_E \leq \bar{C}\ |\underline{|}|\bar{\varepsilon}|\underline{|}|_E \ .$$

Proof: The left hand inequality follows from

$$|\underline{|}|\bar{\varepsilon}|\underline{|}|_E^2 = \bar{v}^T\bar{A}\bar{v} = \bar{v}^T f = \bar{v}^T\underline{A}\bar{v} = \underline{a}_E(\bar{\varepsilon}, \underline{\varepsilon}) \leq |\underline{|}|\bar{\varepsilon}|\underline{|}|_E\ |\underline{|}|\underline{\varepsilon}|\underline{|}|_E \ .$$

The following approach to the right hand inequality is due to Eisenstat [19]. Since $\underline{A}$ is positive definite, using block Gaussian elimination, $A_4 \underline{v}_2 = f_2$, where

$$A_4 := (A_2 - A_3^T A_1^{-1} A_3)$$

is symmetric and positive definite. Then

$$|||\underline{\varepsilon}|||_E^2 = f_2^T A_4^{-T} f_2 .$$

Similarly, A_2 is positive definite, and

$$|||\bar{\underline{\varepsilon}}|||_E^2 = f_2^T A_2^{-T} f_2 .$$

Since A_4 and A_2 are finite-dimensional,

$$\bar{C}^2 := \max_{v \neq 0} v^T A_4^{-T} v \; / \; v^T A_2^{-T} v < \infty.$$

In particular, let $v = f_2$.

For example, if E has no irregular corners, $\underline{a}(E) = 1$, S_E contains bilinears, and $\bar{S}_E - S_E$ contains the biquadratic basis functions

$$(x_E+h_E-x)\,(y_E+h_E-y)\,(y-y_E) \; / \; h_E^3 ,$$

$$(x-x_E)\,(y_E+h_E-y)\,(y-y_E) \; / \; h_E^3 ,$$

$$(x_E+h_E-x)\,(x-x_E)\,(y_E+h_E-y) \; / \; h_E^3 ,$$

$$(x_E+h_E-x)\,(x-x_E)\,(y-y_E) \; / \; h_E^3 ,$$

then

$$A_1 = \begin{bmatrix} 6 & 0 & 0 & 0 \\ 0 & 4 & -1 & -2 \\ 0 & -1 & 4 & -1 \\ 0 & -2 & -1 & 4 \end{bmatrix} / 6 , \quad A_3 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 1 & -1 & -1 \\ -1 & 1 & 1 & -1 \\ -1 & -1 & 1 & 1 \end{bmatrix} / 12 ,$$

$$A_2 = \begin{bmatrix} 13 & 0 & 2 & 0 \\ 0 & 13 & 0 & 2 \\ 2 & 0 & 13 & 0 \\ 0 & 2 & 0 & 13 \end{bmatrix} / 90 ,$$

and $\bar{C} = (11/6)^{1/2} \tilde{=} 1.354$.

4.6.4 Using the residuals and jumps in normal derivatives

We may choose not to solve any linear systems.

By Green's theorem, for v in $\bar{S}_E - S_E$,

$$a_E(U,v) = (LU,v)_E + \langle a\partial U/\partial n, v\rangle_{\partial E} ,$$

so

$$\underline{a}_E(\bar{\varepsilon}_E, v) = (f-LU,v)_E + \langle [a\partial U/\partial n] - a\partial U/\partial n, v\rangle_{\partial E} .$$

In particular,

$$|||\bar{\varepsilon}_E|||_E^2 = (f-LU,\bar{\varepsilon}_E)_E + \langle [a\partial U/\partial n] - a\partial U/\partial n, \bar{\varepsilon}_E\rangle_{\partial E} ,$$

and so

$$|||\bar{\varepsilon}_E|||_E \le C_1(h_E,\underline{a}(E))\ ||f-LU||_E + \qquad (4.21)$$

$$C_2(h_E,\underline{a}(E))\ |[a\partial U/\partial n]-a\partial U/\partial n|_{\partial E}\ ,$$

where

$$C_1(h_E,\underline{a}(E)) := \sup_{v\in S+} ||v||_E/|||v|||_E\ ,$$

$$C_2(h_E,\underline{a}(E)) := \sup_{v\in S+} |v|_{\partial E}/|||v|||_E\ ,$$

and $S+ := \{v : v = v_0(x,y)+C,\ v_0 \text{ in } \bar{S}_E-S_E\ ,\ v \neq 0,\ \text{and } (v,1)_E = 0\}$.

For example, if E has no irregular corners, S_E contains bilinears, and $\bar{S}_E-S_E$ contains the biquadratic basis functions on page 56, then S+ is the span of

$$h_E^{-3}\ (x_E+h_E-x)\ (y_E+h_E-y)\ (y-y_E) - 1/12,$$

$$h_E^{-3}\ (x-x_E)\ (y_E+h_E-y)\ (y-y_E) - 1/12,$$

$$h_E^{-3}\ (x_E+h_E-x)\ (x-x_E)\ (y_E+h_E-y) - 1/12,$$

$$h_E^{-3}\ (x_E+h_E-x)\ (x-x_E)\ (y-y_E) - 1/12,$$

and

$$C_1(h_E,\underline{a}(E)) = h_E\ /\ (22\ \underline{a}(E))^{1/2},$$

$$C_2(h_E,\underline{a}(E)) = \{\ 3\ h_E\ /\ (11\ \underline{a}(E))\ \}^{1/2}.$$

Babuška and Rheinboldt [9] suggest error indicators when S contains

piecewise bilinear basis functions which amount to approximating

$$|||e|||_E^2 \tilde{=} h_E^2/(24\underline{a}(E))\ ||f-LU||_E^2 + \qquad (4.22)$$

$$h_E/(6\underline{a}(E))\ |[a\partial U/\partial n]-a\partial U/\partial n|_{\partial E}^2$$

for interior elements. Since $1/22 \tilde{=} 1.1 \times 1/24$ and $3/11 \tilde{=} 1.6 \times 1/6$, the two schemes are in rough agreement.

If the mesh is uniform, S contains piecewise bilinear basis functions, $\bar{S}$ contains piecewise biquadratic basis functions, u is sufficiently smooth, and h goes to zero, then combining Corollary 4.2, Theorem 4.8, Theorem 4.9, and (4.21),

$$|||e|||^2 \leq (1-\bar{\gamma})^{-2} \sum_E \{(12\underline{a}(E))^{-1/2} h_E ||f-LU||_E +$$

$$(2\underline{a}(E))^{-1/2} h_E^{1/2} |[a\partial U/\partial n]-a\partial U/\partial n|_{\partial E}\}^2,$$

where $\bar{\gamma}$ goes to zero. Similar results hold for higher-order elements and nonuniform grids.

In numerical tests, we have observed that quantities of the form

$$\sum_E C_1(h_E,\underline{a}(E))\ ||f-LU||_E^2 + C_2(h_E,\underline{a}(E))\ |[a\partial U/\partial n]-a\partial U/\partial n|_{\partial E}^2$$

or

$$\sum_E \{C_1(h_E,\underline{a}(E))\ ||f-LU||_E + C_2(h_E,\underline{a}(E))\ |[a\partial U/\partial n]-a\partial U/\partial n|_{\partial E}\}^2$$

only seem to approach $|||e|||^2$ to within a multiplicative constant

depending on u. In Chapter 7, we present numerical results using the error estimator

$$\underline{\underline{\varepsilon}} := \left(\sum_E \underline{\underline{\varepsilon}}_E^2 \right)^{1/2},$$

where

$$\underline{\underline{\varepsilon}}_E^2 := \underline{\underline{K}}(k_E)\{h_E^2/\underline{a}(E)\ \|f-LU\|_E^2 + 4h_E/\underline{a}(E)\ \|[a\partial U/\partial n]-a\partial U/\partial n|_{\partial E}^2\}, \quad (4.23)$$

and

$$\underline{\underline{K}}(k_E) := (k_E(k_E+1)(k_E+2))^{-1}. \quad (4.24)$$

$\underline{\underline{\varepsilon}}_E$ is chosen to extend formula (4.22), and to perform well on the problem

$$-D_x^2 u - D_y^2 u = f(x,y) \text{ in } \Omega \quad , \quad u = 0 \text{ on } \partial\Omega,$$

when $u = \sin(\pi x)\sin(\pi y)$, k_E and h_E are uniform, and h goes to zero.

4.7 Extensions

If $b(x,y) = 0$ in Ω, in order to insure nonsingular linear systems, we impose the Neumann consistency conditions $(u,1) = 0$, $(U,1) = 0$, and $(\tilde{\varepsilon},1)_E = 0$. We still have

$$|||v|||_E \leq C\ \|v\|_{1,E} \quad \text{for all } v \text{ in } H^1(E),$$

and globally,

$$\|e\|_1 \leq C\ |||e|||.$$

Since $(\tilde{e}_E,1)_E = 0$ implies $\|\tilde{e}_E\|_E \le C\, h_E\, \|\tilde{e}_E\|_{1,E}$,

$$\|\tilde{e}\|_{1,E} \le C\, |||\tilde{e}|||_E$$

for h_E sufficiently small. (Similarly, $\|\tilde{e}\|_{1,E} \le C\, |||\tilde{e}|||_E$.) Thus, Theorem 4.6 and Theorem 4.7 hold for h sufficiently small.

If $a_E(\cdot,\cdot)$ is of the form

$$a_E(v,w) = (a(x,y)\, D_x v, D_x w)_E + (a(x,y)\, D_y v, D_y w)_E +$$

$$(b(x,y)\, v, w)_E + (c(x,y)\, D_x v, w)_E + (d(x,y)\, D_y v, w)_E ,$$

then, under suitable assumptions, $\|\tilde{e}\|^2_{1,E} \le C\, a_E(\tilde{e},\tilde{e})$ and $\|\tilde{e}\|^2_{1,E} \le C\, a_E(\tilde{e},\tilde{e})$ as h goes to zero (see Schatz [27]). Thus, Theorem 4.6 and Theorem 4.7 hold for h sufficiently small.

For problems with non-homogeneous Neumann boundary conditions

$$a\partial u/\partial n = g(x,y) \quad \text{on } \partial\Omega\ ,\ g \text{ in } H^0(\partial\Omega),$$

if we set

$$a(u,v) = (f,v) + \langle g,v\rangle \quad \text{for all } v \text{ in } H^1(\Omega),$$

$$a(U,v) = (f,v) + \langle g,v\rangle \quad \text{for all } v \text{ in } S,$$

and

$$[a\partial U/\partial n] := g(x,y) \text{ if s is on } \partial\Omega,$$

then we obtain the same results as before.

For problems with homogeneous Dirichlet boundary conditions

$$u = 0 \quad \text{on } \partial\Omega,$$

the solution u in $H_0^1(\Omega)$ satisfies

$$a(u,v) = (f,v) \quad \text{for all } v \text{ in } H_0^1(\Omega).$$

Similarly, the equations for U, $\tilde{\varepsilon}$, $\underline{\varepsilon}$, and $\bar{\varepsilon}$ are now posed on subspaces of $H_0^1(\Omega)$.* Since all the functions of interest now vanish on $\partial\Omega$, the other results follow as before. Problems with non-homogeneous Dirichlet boundary conditions which are <u>exactly satisfied</u> are handled similarly.

When Dirichlet boundary conditions are not exactly satisfied by U, in element E touching $\partial\Omega$, we define $\tilde{\varepsilon}_E$ in $\bar{S}_E$ by the equations

$$a_E(\tilde{\varepsilon}_E, v) = \tilde{F}_E(v - I_E(v)) \text{ for all } v \text{ in } \bar{S}_E \text{ which vanish on } \partial\Omega,$$

$$\tilde{\varepsilon}_E = \varepsilon \text{ on } \partial\Omega.$$

We define $\tilde{\varepsilon}$ from $\tilde{\varepsilon}_E$ as before. Let $\tilde{\varepsilon}_b$ be any function in $\tilde{S}$ such that

---

*Equation (4.19) is no longer needed in elements touching $\partial\Omega$.

$\tilde{\varepsilon}_0 := \tilde{\varepsilon} - \tilde{\varepsilon}_b$ vanishes on $\partial\Omega$, and let $\varepsilon_0 := \varepsilon - \tilde{\varepsilon}_b$. By the same proof as for Theorem 4.1,

$$|||\varepsilon_0||| \leq |||\tilde{\varepsilon}_0|||,$$

so

$$(1-\bar{\gamma})\ |||e||| \leq |||\tilde{\varepsilon}_0||| + |||\tilde{\varepsilon}_b|||.$$

Similarly, under the assumptions of Theorem 4.6,

$$|||\tilde{\varepsilon}|||_E \leq C \sum_{E' \in N(E)} \{\ ||e||^2_{1,E'} + h^2_{E'}\ ||e||^2_{2,E'}\ \}^{1/2} + |||\tilde{\varepsilon}_b|||_E\ ,$$

and under the assumptions of Theorem 4.7,

$$|||\tilde{\varepsilon}||| \leq \tilde{C}\ |||e||| + |||\tilde{\varepsilon}_b|||.$$

If

$$\tilde{\varepsilon} = \sum \tilde{\varepsilon}_I\ B_I\ ,$$

where $\{B_I\}$ is the basis for $\bar{S}$, we choose

$$\tilde{\varepsilon}_b = \sum_{B_I \in \bar{S}_b} \tilde{\varepsilon}_I\ B_I\ ,$$

where

$$\bar{S}_b := \text{span}\{B_I : B_I \text{ in } \bar{S},\ B_I(x,y) \neq 0,\ (x,y) \text{ in } \partial\Omega\}.$$

Since the area of $\text{supp}(\tilde{\varepsilon}_b)$ is $O(h)$, we expect $|||\tilde{\varepsilon}_b||| = o(|||\tilde{\varepsilon}|||)$ as h

goes to zero. Since e is nonzero on $\partial\Omega$, $\tilde{\varepsilon}$ should be in $\bar{S}_E$ rather than $\bar{S}_E - S_E$ when E touches $\partial\Omega$.

Analogous results hold for $\underline{\varepsilon}$ and $\bar{\varepsilon}$.

As in Babuška and Rheinboldt [10], on an element side s of E which touches $\partial\Omega$, we replace the term

$$\underline{\underline{K}}(k_E)\ 4h_E/\underline{a}(E)\ |[a\partial U/\partial n]-a\partial U/\partial n|_s^2$$

with the term

$$3\underline{a}(E)/h_E\ |e|_s^2$$

in formula (4.23).

4.8 Summary

The error indicators discussed in this chapter are:

$|||\tilde{\varepsilon}_E|||_E$, where $\tilde{\varepsilon}_E$ is in $\bar{S}_E$ and

$$a_E(\tilde{\varepsilon}_E, v) = \begin{cases} \tilde{F}_E(v) & \text{for all } v \text{ in } \bar{S}_E - S_E \\ 0 & \text{for all } v \text{ in } S_E \ , \end{cases}$$

$|\underline{|}\underline{\varepsilon}_E|\underline{|}_E$, where $\underline{\varepsilon}_E$ is in $\bar{S}_E$, $(\underline{\varepsilon}_E, 1)_E = 0$, and

$$\underline{a}_E(\underline{\varepsilon}_E, v) = \begin{cases} \tilde{F}_E(v) & \text{for all } v \text{ in } \bar{S}_E - S_E \\ 0 & \text{for all } v \text{ in } S_E \ , \end{cases}$$

$|||\bar{\varepsilon}_E|||_E$, where $\bar{\varepsilon}_E$ is in $\bar{S}_E - S_E$ and

$$\underline{a}_E(\bar{\varepsilon}_E, v) = \tilde{F}_E(v) \quad \text{for all } v \text{ in } \bar{S}_E - S_E ,$$

and $\underline{\underline{\varepsilon}}_E$, where

$$\underline{\underline{\varepsilon}}_E^2 = \underline{\underline{K}}(k_E) \{ h_E^2/\underline{a}(E) \, \|f-LU\|_E^2 + 4h_E/\underline{a}(E) \, |[a\partial U/\partial n] - a\partial U/\partial n|_{\partial E}^2 \} .$$

The corresponding <u>error estimators</u> are

$$|||\tilde{\varepsilon}||| = (\sum_E |||\tilde{\varepsilon}_E|||_E^2)^{1/2},$$

$$|||\underline{\underline{\varepsilon}}||| = (\sum_E |||\underline{\underline{\varepsilon}}_E|||_E^2)^{1/2},$$

$$|||\bar{\varepsilon}||| = (\sum_E |||\bar{\varepsilon}_E|||_E^2)^{1/2},$$

and

$$\underline{\underline{\varepsilon}} := (\sum_E \underline{\underline{\varepsilon}}_E^2)^{1/2}.$$

We define the corresponding <u>relative errors</u> in the estimators*

---

*Babuška and Rheinboldt call (the relative error + 1) the "efficiency index" [10] or the "effectivity index" [9].

$$\tilde{\rho} := (|||\tilde{\varepsilon}||| - |||e|||) / |||e|||,$$

$$\underline{\rho} := (|\underline{|}\underline{\varepsilon}|\underline{|}| - |||e|||) / |||e|||,$$

$$\bar{\rho} := (|\underline{|}\bar{\varepsilon}|\underline{|}| - |||e|||) / |||e|||,$$

and

$$\underline{\rho} := (\underline{\varepsilon} - |||e|||) / |||e|||.$$

We have shown, for sufficiently smooth u, that as h goes to zero,

$$(1-\bar{\gamma})\ |||e||| \leq |||\tilde{\varepsilon}||| \leq \tilde{C}\ |||e|||,$$

$$(1-\bar{\gamma})\ |||e||| \leq |\underline{|}\underline{\varepsilon}|\underline{|}| \leq \tilde{C}\ |||e|||,$$

$$(1-\bar{\gamma})\ \bar{C}^{-1}\ |||e||| \leq |\underline{|}\bar{\varepsilon}|\underline{|}| \leq \tilde{C}\ |||e|||,$$

and

$$\underline{C}\ |||e||| \leq \underline{\varepsilon},$$

where the positive constants $\tilde{C}$, $\bar{C}$, and $\underline{C}$ are independent of h, and if u is sufficiently smooth, $\bar{\gamma}$ goes to zero if h goes to zero.

CHAPTER 5

Heuristic refinement criteria

5.1 Introduction

The guiding principle in choosing a finite element space is to minimize the norm of the error as a function of work. However, there is not usually enough information available to choose an appropriate space a priori. Instead, it is necessary to choose a sequence of spaces adaptively. Each succeeding space in the sequence is chosen on the basis of criteria computable from the preceding space, with the criteria tending to satisfy the guiding principle when a heuristic model of error behavior holds.

When the polynomial orders of the basis functions are uniform, the resulting refinement algorithms tend to approximately equalize the sizes of the errors in the elements in the next space. In finite element methods, for example, this heuristic refinement criterion is used in Fried and Yang [20], Babuška and Rheinboldt [8], and Bank and Sherman [13]. When the polynomial orders of the basis functions vary locally, the heuristic refinement criteria must be somewhat more complicated (e.g., Brandt [16]).

In Sections 5.2 - 5.5, we develop two heuristic models of error behavior for local-mesh, local-order finite element spaces, and an adaptive refinement algorithm which uses the models. In Section 5.6, we discuss drawbacks inherent in the models. In Section 5.7, we present the asymptotic expected error and work behavior for three problem classes consisting of problems with:

1. smooth solutions.
2. solutions with point singularities.
3. solutions with line singularities.

Except in the presence of line singularities, with appropriate local mesh and local order refinement, the expected error converges to zero faster than inverse polynomially with respect to work.

5.2 A model of work

The guiding principle in choosing a "good" finite element space S can be stated as

$$\text{minimize } |||e||| \text{ as a function of } W, \tag{5.1}$$

where $|||e|||$ is the error in the energy norm, and W is the work. However, there is not usually enough information available to choose an appropriate space a priori. Thus, we must choose a sequence of nested spaces

$$S_1 \text{ contained in } S_2 \text{ contained in } \ldots S_n ,$$

with finite element solutions $U_1,\ldots,U_n$, taking $S = S_n$ (for instance, because the work limit is reached). We choose each S_{i+1} <u>adaptively</u>, on the basis of information obtained from S_i. Since the spaces are

nested, the errors $\{e_i\}$ and work counts $\{W_i\}$ satisfy

$$|||e_1||| \geq |||e_2||| \geq \ldots \geq |||e_n|||$$

and

$$W_1 < W_2 < \ldots < W_n .$$

The work counts can represent either storage or computer time. If the work counts represents storage, $W = W_n$. If the work counts represents time, $W = \sum_{i=1}^{n} W_i$, and if

$$W_i \leq \gamma W_{i+1} , \quad 0 \leq \gamma < 1 , \quad 1 \leq i \leq n, \tag{5.2}$$

then $W \leq (1-\gamma)^{-1} W_n$. Thus, W is usually proportional to W_n . Since it is possible that $n = i+1$, in the remainder of this chapter, we take as the "local guiding principle" a local version of (5.1):

$$\text{given } S_i \text{ , choose } S_{i+1} \text{ to minimize } |||e_{i+1}||| \tag{5.3}$$

$$\text{as a function of } W_{i+1}$$

(subject to the possible enforcement of (5.2)).

Let S be the current finite element space S_i in the sequence, with mesh M, unrefined elements {E}, and error e. Let $z := (x,y)$ denote a point in Ω. Let $h(z)$, $p(z)$, and $w(p(z))$ be piecewise constant functions defined by

$h(z) := h_E$ = the local mesh size,

$p(z) := p_E := k_E - 1$ = the local polynomial degree,

$w(p(z)) := w_E$ = the local work count,

if z is in element E. We make the heuristic assumption that the work W for S is

$$W \tilde{=} \sum_E w_E \,. \tag{5.4}$$

This is reasonable if W is the dimension of S, or (since the mesh is 1-irregular) the storage for the stiffness matrix for S, or the number of arithmetic operations needed to solve the finite element system for S when the work is proportional to the number of elements in the mesh (see Section 6.3). It is <u>not</u> reasonable if the work grows faster than the number of elements: then, W depends on S globally as well as locally.

5.3 A model of predicted error behavior

In this section, we construct a model of predicted error behavior based on standard a priori error estimates, and derive heuristic refinement criteria which tend to satisfy (5.3) when the model holds.

Let $m(z)$ and $\tilde{p}(z)$ be piecewise constant functions defined by

$$m(z) := \max\{ k : \|u\|_{k,E} < \infty \} ,$$

$$\tilde{p}(z) := \min\{p(z), m(z)-1\} ,$$

if z is in element E. m indicates the maximum local smoothness of u in E, and $\tilde{p}$ indicates the expected convergence rate in the energy norm using polynomials of degree p (order k_E). Standard a priori error estimates (e.g., Ciarlet [17], Theorem 3.1.5 and proofs of Theorems 3.2.1 and 3.2.2) yield

$$|||e|||^2 \leq \iint_\Omega C(\tilde{p}(z))\ a(z)\ G(z,\tilde{p}(z))\ h(z)^{2\tilde{p}(z)}\ dz, \tag{5.5}$$

where a(z) is the coefficient in (4.1), and

$$G(z,\tilde{p}) := \sum_{j=0}^{\tilde{p}+1} \left(D_x^j\ D_y^{\tilde{p}+1-j}\ u(z) \right)^2$$

is in $L^2(\Omega)$ by the definition of $\tilde{p}$.

Following Brandt [16], Chapter 8, we make the crucial assumption that, for heuristic purposes, (5.5) holds in <u>each</u> element, with <u>equality</u>:

$$|||e|||_E^2 \tilde{=} \xi_E := \iint_E C(\tilde{p}(z))\ a(z)\ G(z,\tilde{p}(z))\ h(z)^{2\tilde{p}(z)}\ dz, \tag{5.6}$$

so that

$$|||e|||^2 \tilde{=} \xi := \sum_E \xi_E .$$

We also assume, for heuristic purposes, that

a and G are approximately constant in each E (5.7)

and

C, G, and W are (formally) differentiable in h and p. (5.8)

Assumptions (5.4), (5.6), (5.7), and (5.8) comprise our first heuristic model of predicted error behavior. The form of (5.6) is important, but not the specific values of $C(\tilde{p})$ and $G(z,\tilde{p})$. We make assumption (5.8) only for convenience: all derivatives with respect to h and p can be replaced by divided differences.

In the following paragraphs, using (5.4), (5.6), (5.7), and (5.8), we estimate the marginal efficiency of refinement in each element E, obtaining

$$\lambda_E^{(h)} \tilde{=} - (\partial\xi/\partial h_E) \; / \; (\partial W/\partial h_E) \tag{5.9}$$

for mesh refinement (h-refinement), and

$$\lambda_E^{(p)} \tilde{=} - (\partial\xi/\partial p_E) \; / \; (\partial W/\partial p_E) \tag{5.10}$$

for order refinement (p-refinement). Let

$$\lambda_E := \max\{\lambda_E^{(h)}, \lambda_E^{(p)}\}.$$

We use these quantities in the following adaptive refinement algorithm:

1. Find the unrefined element E with the largest λ_E .

2. Determine the refinement cutoff $\lambda_{CUT} < \lambda_E$ (e.g., as in Section 5.5).

3. For each unrefined element E with $\lambda_E \geq \lambda_{CUT}$:

   a. If $\lambda_E^{(h)} \geq \lambda_E^{(p)}$, refine the mesh in E.

   b. If $\lambda_E^{(h)} < \lambda_E^{(p)}$, refine the order in E.

4. Take S_{i+1} to be the resulting finite element space.

We now determine $\lambda_E^{(h)}$ and $\lambda_E^{(p)}$. First, we estimate $\tilde{p}$ in each unrefined element E. Suppose the order in E has never been refined, and (from a previous finite element space in the sequence) we have an approximation $\xi_{f(E)}$ to the square of the error in f(E) before f(E) was refined. By (5.6) and (5.7),

$$\xi_{f(E)} \tilde{=} \xi_E \, 2^{2(\tilde{p}+1)},$$

so we estimate

$$\tilde{p} \tilde{=} (\ln(\xi_{f(E)}) - \ln(\xi_E)) / \ln(4) - 1.$$

On the other hand, suppose the order in E has been refined, and we have an approximation $\bar{\xi}_E$ to the square of the error in element E with k_E one less than its current value. If

$$\xi_E \; / \; \bar{\xi}_E \leq p_{CUT} \tag{5.11}$$

for some constant $p_{CUT} < 1$ (e.g., as in Section 5.4), there may still be higher derivatives of u in E which can be exploited, so we estimate $\tilde{p} \tilde{=} p_E$. Otherwise, we retain our previous estimate of $\tilde{p}$. In order for our estimate of $\tilde{p}$ to be logically consistent, we restrict it to the interval $(0,p_E]$.

Using our estimate of $\tilde{p}$, by (5.7) and (5.9), we estimate

$$\begin{aligned} \lambda_E^{(h)} &\tilde{=} - \{\partial(h_E^2 \; a \; C \; G \; h^{2\tilde{p}})/\partial h\} \; / \; \{\partial(h_E^2 \; w \; h^{-2})/\partial h\} \qquad (5.12) \\ &= - \{2 \; \tilde{p} \; (a \; C \; G \; h^{2\tilde{p}}) \; h^{-1}\} \; / \; \{(-2) \; w \; h^{-3}\} \\ &= (\; \tilde{p} \; / \; w \;) \; \xi_E \end{aligned}$$

in element E.

There are two cases to consider in estimating $\lambda_E^{(p)}$. If $p \geq m-1$, refining the order will not change the error at all (in (5.6)). Thus, $\lambda_E^{(p)} = 0$. On the other hand, if $p < m-1$, by (5.7) and (5.10),

$$\lambda_E^{(p)} = - \{\partial(\xi_E)/\partial p\} \; / \; w',$$

where $w' := \partial w/\partial p$. If $u(z) \tilde{=} C \sin(\theta x) \sin(\theta y)$ in E, where θ is the approximate "frequency" of u in E, then

$$||u||^2_{p+1,E} \tilde{=} C_1 \; \theta^{2(p+1)},$$

and by (5.6), since $C(\tilde{p}) \tilde{=} (p!)^{-2}$ for Taylor's theorem,

$$\xi_E \tilde{=} C_2\ (p!)^{-2}\ h^2\ \theta^{2(p+1)}\ h^{2p} \tag{5.13}$$

$$= C_2\ (\theta\ h)^{2(p+1)}\ (p!)^{-2}.$$

Since θ is independent of h and p, by (5.13),

$$\theta \tilde{=} (p\ \xi_E\ /\ \bar{\xi}_E)^{1/2}\ h^{-1}. \tag{5.14}$$

Using the crude bound $\partial(p!)/dp \leq p\ p!/2$,

$$\partial(\xi_E)/\partial p \tilde{=} \xi_E\ (2\ \ln(\theta\ h) - p),$$

so we estimate

$$\lambda_E^{(p)} \tilde{=} \begin{cases} \xi_E\ (-2\ \ln(\theta\ h) + p)\ /\ w' & \text{if } p < m-1 \\ 0 & \text{if } p \geq m-1. \end{cases} \tag{5.15}$$

In practice, we estimate ξ_E by the squared error indicator $|||\tilde{\varepsilon}|||_E^2$ (or $|||\underline{\varepsilon}|||_E^2$, $|||\bar{\varepsilon}|||_E^2$, or ε_E^2) in formulae (5.12), (5.14), (5.15), and the formulae for $\tilde{p}_E$.

Expressions (5.12) and (5.15) indicate that λ_E tends to decrease as E is refined. Thus, the adaptive refinement algorithm tends to approximately equalize $\{\lambda_E\}$. When the order is uniform, this reduces to "asymptotic equidistribution" of the errors among the elements (e.g., Fried and Yang [20]).

5.4 A symmetrized model of predicted error behavior

In this section, we construct a second model of predicted error behavior based on symmetrized a priori error estimates.

In considering combined h and p refinement, Babuška and Dorr [2] present an a priori error estimate of the form

$$|||e|||^2 \leq C(kmax) \iint_\Omega a(z)\ G(z,m(z))\ h(z)^{2\tilde{p}(z)} p(z)^{-2(m(z)-1)} dz. \quad (5.16)$$

Since C and G are independent of p, this estimate is more symmetric in h and p than (5.5). Moreover, it indicates the manner in which convergence occurs when $p > m-1$. Theorems and numerical evidence in [11] show the exponent in p is optimal in the presence of singularities of u not located at mesh nodes, and that if all the singularities of u are located at mesh nodes, the exponent of p can double. The effect of this doubling will be considered at the end of this section.

We make the heuristic assumption that (5.16) holds in <u>each</u> element, with <u>equality</u>:

$$|||e|||_E^2 \tilde{=} \xi_E \quad (5.17)$$

$$:= C(kmax) \iint_E a(z)\ G(z,m(z))\ h(z)^{2\tilde{p}(z)}\ p(z)^{-2(m(z)-1)}\ dz.$$

Assumptions (5.4), (5.7), (5.8), and (5.17) comprise our second heuristic model of predicted error behavior. The resulting estimates of $\{\lambda_E^{(h)}\}$, $\{\lambda_E^{(p)}\}$, and $\{\lambda_E\}$ can be used in the refinement algorithm in Section 5.3.

In our second model, if the order in unrefined element E has never been refined, then by (5.17) and (5.7), we estimate

$$\tilde{p} \;\tilde{=}\; (\ln(\xi_{f(E)}) - \ln(\xi_E)) \;/\; \ln(4) - 1.$$

On the other hand, if the order in E has been refined, then if $\tilde{p} < m-1$,

$$\xi_E \;/\; \xi_E^- \;\tilde{=}\; h^2 \; (p/(p-1))^{-2(m-1)},$$

else if $\tilde{p} \geq m-1$,

$$\xi_E \;/\; \xi_E^- \;\tilde{=}\; (p/(p-1))^{-2(m-1)}.$$

Since the function $(p/(p-1))^{-2(p-1)}$ equals 1/4 at 2, and is monotone decreasing with limit e^{-2} as k goes to ∞, we decide whether to increase our estimate of $\tilde{p}$ on the basis of (5.11), with $p_{CUT} = e^{-2}$. In order for our estimate of $\tilde{p}$ to be logically consistent, we restrict it to the interval $(0,p_E]$.

As in Section 5.3, we estimate

$$\lambda_E^{(h)} \;\tilde{=}\; (\tilde{p} \;/\; w) \; \xi_E \; . \tag{5.18}$$

If $\tilde{p} < m-1$, we estimate

$$\lambda_E^{(p)} \;\tilde{=}\; \xi_E \; (-2 \ln(h) + 2 \; (m-1)/p) \;/\; w'. \tag{5.19}$$

Otherwise, we estimate

$$\lambda_E^{(p)} \;\tilde{=}\; \xi_E \; (2 \; (m-1)/p) \;/\; w'. \tag{5.20}$$

In practice, we approximate ξ_E by the squared error indicator $|||\tilde{\varepsilon}|||_E^2$ (or $|||\underline{\varepsilon}|||_E^2$, $|||\bar{\underline{\varepsilon}}|||_E^2$, or $\underline{\varepsilon}_E^2$) in formulae (5.18), (5.19), (5.20), and the formulae for $\tilde{p}$.

Suppose the exponent of p in (5.17) is doubled. Then our estimate of $\lambda_E^{(h)}$ is exactly as before, and our estimate of $\lambda_E^{(p)}$ changes by a factor of 2 when $p \geq m-1$, and by a small factor when $p < m-1$ and h_E is small. The main effect is to encourage order refinement when m is small and $p > m-1$. But then $\lambda_E^{(p)}$ is usually much smaller than $\lambda_E^{(h)}$, so the practical effect on refinement behavior is slight. In general, the form of (5.17) is more important than the values of the exponents.

5.5 Choosing the refinement cutoff

A simple choice for λ_{CUT} is

$$\lambda_{CUT} = C\,\lambda_{\mathbf{E}}$$

for some constant $0 < C < 1$, where

$$\lambda_{\mathbf{E}} = \max_E\{\lambda_E\}.$$

If λ_{CUT} is too large, (5.2) may not hold. If λ_{CUT} is too small, we may refine an element with small λ_E when it may be more efficient to refine an element with large λ_E twice. We can avoid this situation by letting λ_{CUT} approximate the largest λ_E for a once-refined element. This quantity can be estimated by the marginal efficiency of refinement for element E after it has been refined once:

If $\lambda_E = \lambda_E^{(h)}$, and the order in E has never been refined, we can estimate

$$\lambda_{CUT} \tilde{=} (\tilde{p}_E / w_E) \xi_E^2 / \xi_{f(E)} .$$

If $\lambda_E = \lambda_E^{(h)}$ and the order in E has been refined, then by (5.13) and (5.12) (or (5.17) in our second model), we can estimate

$$\lambda_{CUT} \tilde{=} 4^{-(\tilde{p}_E+1)} \lambda_E^{(h)}.$$

If $\lambda_E = \lambda_E^{(p)}$, then in our first model, by (5.13) and (5.15), we can estimate

$$\lambda_{CUT} \tilde{=} \lambda_E^{(p)} (\theta_E h_E)^2 / (p_E+1) (w'(\tilde{p}_E+1)/w'(\tilde{p}_E+1)),$$

and in our second model, by (5.17)-(5.20), we can estimate

$$\lambda_{CUT}^{EST} \tilde{=} \lambda_E^{(p)} h_E^2 ((p_E+1)/p_E)^{-2(m_E-1)} (w'(\tilde{p}_E)/w'(\tilde{p}_E+1)).$$

5.6 Drawbacks

The heuristic assumptions leading to our estimates of $\lambda_E^{(p)}$ may be too crude to yield useful refinement information. We base an alternate refinement algorithm solely on $\lambda_E^{(h)}$ and $\tilde{p}$:

1. Find the unrefined element E with the largest $\lambda_E^{(h)}$.

2. Determine $\lambda_{CUT} < \lambda_E^{(h)}$.

3. For each unrefined element E with $\lambda_E^{(h)} \geq \lambda_{CUT}$:

   a. If $p > \tilde{p}$, refine the mesh in E.

   b. If $p \leq \tilde{p}$, refine the order in E.

4. Take S_{i+1} to be the resulting finite element space.

If the order in element E is refined, $\tilde{p}$ is estimated by (5.11). Then if ξ_E^- and ξ_E behave similarly to $|||e(p-1)|||_E^2$ and $|||e|||_E^2$, p tends to be one more than its (usually) optimal value m-1. Conversely, if $\xi_E^- << |||e(p-1)|||_E^2$, then p tends to be less than m-1. The latter situation arises most frequently when $k_s > k_E$ on the sides of E. Thus, in our current code, if $k_s \geq k_E+1$ on any three sides of E, or $k_s \geq k_E+2$ on any side of E, then the next time E is refined, the order is refined, regardless of the size of $\tilde{p}$.

Lack of de-refinement penalizes early order refinement. For example, suppose $h \leq 1/4$ and $k \geq 3$ are desired near the origin. If the order is refined first, then $k \geq 3$ everywhere (Figure 5-1).

The resulting finite element space may not satisfy (5.3), even if the heuristic models of error behavior hold. For instance, if $\lambda_E^{(h)} < \lambda_E^{(p)}$, it may still be more efficient to refine E in h if the resulting decrease in ξ_E is greater with mesh refinement than with order

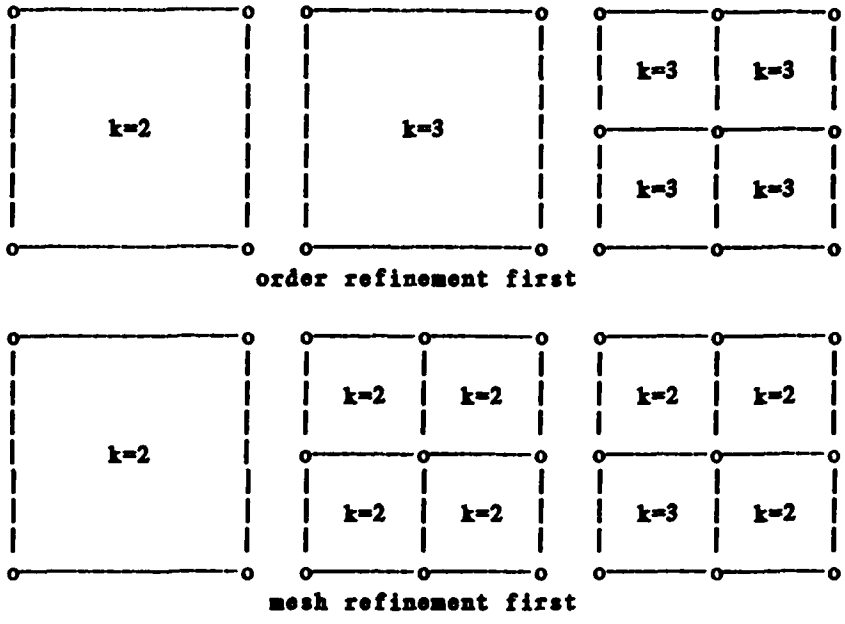


Figure 5-1: Early order refinement penalty

refinement, and if the other λ_E's are very small.

Rigorous models are preferable to the heuristic models used in this chapter. A rigorous approach for <u>one</u>-dimensional problems is developed in Babuška and Rheinboldt [7].

5.7 Expected error behavior

In this section, we model the asymptotic behavior of error with respect to work for three problem classes consisting of problems with:

1. smooth solutions.
2. solutions with point singularities.
3. solutions with line singularities.

We use the heuristic model from Section 5.3. We expect similar behavior using the heuristic model from Section 5.4.

5.7.1 A smooth solution

Suppose u is smooth everywhere in Ω, and h(z) and p(z) are uniform. By assumption (5.6),

$$\xi \tilde{=} C(p)h^{2p},$$

and by assumption (5.4),

$$W \tilde{=} p^{\beta} h^{-2}, \beta \geq 2.$$

Minimizing ξ while holding W fixed, $2\ p/h\ \xi - \lambda\ 2/h\ W = 0$ and $((C'/C) + 2\ \ln(h))\xi + \lambda\ \beta/p\ W = 0$ imply

$$h \tilde{=} e^{-\gamma/2}$$

and

$$\xi \tilde{=} C\ e^{-\gamma p},$$

where

$$\gamma := \beta + C'/C.$$

Thus, it is best to fix the mesh and only refine the order. Suppose

$$C(p) = A^{p}\ ,\ A > 1.$$

Since $C'/C = \ln(A)$, then $A\ e^{-\gamma} = A\ e^{-\beta-\ln(A)} < 1$, so

$$\xi \tilde{=} (A\ e^{-\gamma})^{p}$$

is exponentially decreasing in p. But $W \tilde{=} p^{\beta}\ e^{\gamma}$ is polynomial in p, so if there exists $A < \infty$ such that $C(p) \lesssim A^{p}$ for sufficiently large p, then ξ is exponentially decreasing with respect to W.

5.7.2 A point singularity

Suppose u has an isolated singularity at a point σ in Ω, such that near σ, u has $1+\alpha$ derivatives in L^2, and u is smooth away from σ. Suppose there are unrefined elements at levels 1 to L in the mesh for S_L, and at each level only the elements which touch the point σ are originally refined. Suppose p(z) uniformly equals L. Then

$$\xi_E \tilde{=} C(\alpha)\ h_E^{2+2\alpha} \quad \text{if E touches } \sigma$$

and

$$\xi_E \tilde{=} C(p)\ h_E^{2+2p} \quad \text{if E does not touch } \sigma.$$

By the proof of (2.4), there are no more than a constant number of elements at each level, so

$$\xi = \sum_E \xi_E$$

$$\leq C_1(p)\ \{(2^{-L})^{2+2\alpha} + \sum_{i=1}^{L-1} (2^{-i})^{2+p}\}$$

$$\leq C_1(L)\ 2^{-L},$$

and

$$W \leq C_2 \sum_{i=1}^{L} p^{\beta}$$

$$= C_2\ L^{\beta+1}.$$

If $C_1(L) \leq A^L$ for some $A < 2$ as L goes to ∞, then ξ is exponentially decreasing with respect to W. Babuška and Dorr [2] make a similar

conjecture, and under suitable assumptions, prove that convergence is faster than polynomial.

5.7.3 A line singularity

Suppose u has a line singularity at a line σ in Ω, such that near σ, u has $1+\alpha$ derivatives in L^2, and u is smooth away from σ. To fix ideas, suppose p(z) is uniform, σ traverses Ω from top to bottom, and there are elements at levels 1 to L in the mesh for S_L, where at each level the elements which touch the line σ are present in S_L. Then

$$\xi_E \tilde{=} C(\alpha)\ h_E^{2+2\alpha} \quad \text{if E touches } \sigma$$

and

$$\xi_E \tilde{=} C(p)\ h_E^{2+2p} \quad \text{if E does not touch } \sigma.$$

There are at least 2^L elements with level L touching σ, so

$$\xi \geq C_1(\alpha)\ 2^L\ (2^{-L})^{2+2\alpha}$$

and

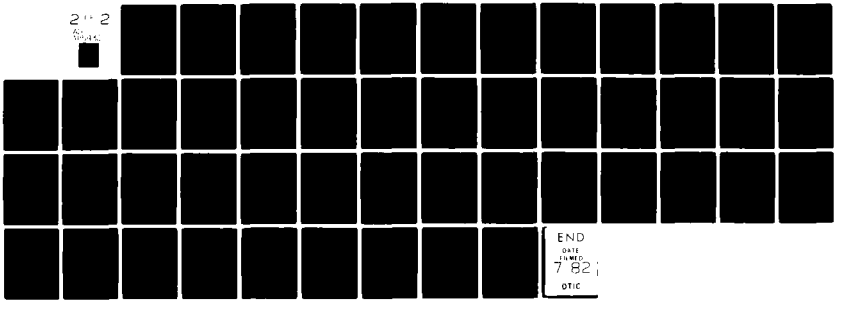
$$W \geq C_2(\alpha)\ 2^L.$$

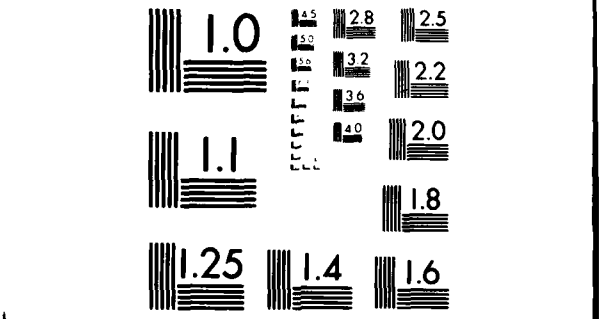
Then

$$\xi \geq C_3(\alpha)\ W^{-1-2\alpha}$$

<u>regardless</u> of how p is chosen. Convergence cannot be faster than polynomial, since the complexity of geometrically approximating σ asymptotically dominates the cost.

AD-A115 432 YALE UNIV NEW HAVEN CT DEPT OF COMPUTER SCIENCE F/G 12/1
LOCAL-MESH, LOCAL-ORDER, ADAPTIVE FINITE ELEMENT METHODS WITH A--ETC(U)
DEC 81 A WEISER N00014-76-C-0277
UNCLASSIFIED TR-213 NL





CHAPTER 6

Computational aspects

6.1 Introduction

In this chapter, we discuss some of the computational aspects of the methods presented in Chapters 2-5. Since we have not yet implemented all the procedures discussed, detailed operation counts are omitted.

In Section 6.2, we consider the efficient assembly of the finite element system (cf. Eisenstat and Schultz [18], Weiser, Eisenstat, and Schultz [31]). In Section 6.3, we investigate the complexity of sparse direct elimination with nested-dissection-type ordering of the unknowns, for the three classes of problems discussed in Section 5.7. The operation counts and storage for the problems with singularities are linear (optimal-order) in the number of elements. In Section 6.4, we consider the efficient computation of the a posteriori error estimators.

We now outline the computational tasks to be performed.

There are four computational steps to perform for each finite element space S in the sequence $\{S_i\}$.

1. Form the linear system $A\,U = f$, where

   $\{B_I\}$ is the basis for S,

   $$U(x,y) = \sum_J U_J B_J(x,y),$$

   $$A_{IJ} = \sum_E A_{IJ}^E, \qquad A_{IJ}^E := a_E(B_I,B_J),$$

   $$f_I = \sum_E f_I^E, \qquad f_I^E := (f,B_I)_E.$$

   After initializing A and f to zero, for each E,

   a. Assemble $\tilde{A}^E$ and $\tilde{f}^E$, the element stiffness matrix and right hand side for the basis $\{\tilde{B}_I|_E\}$, where $\{\tilde{B}_I\}$ is the basis which would result if E had no irregular corners.

   b. Obtain A^E and f^E from $\tilde{A}^E$ and $\tilde{f}^E$, by changing variables from $\{\tilde{B}_I\}$ to $\{B_I\}$.

   c. Add A^E and f^E to A and f.

2. Solve $A\,U = f$.

3. For each E, compute $|||\tilde{\varepsilon}|||_E$, $|||\underline{\varepsilon}|||_E$, $|||\bar{\underline{\varepsilon}}|||_E$, or $\underline{\varepsilon}_E$.

4. Depending on the error indicators computed in Step 3, either stop, or adaptively choose S_{i+1}, and go to Step 1.

Each integral $\iint_E v(x,y)\,dx\,dy$ is approximated by

$$\sum_{s=1}^{G(E)} \sum_{t=1}^{G(E)} w_s\, w_t\, v(x_s,y_t), \tag{6.1}$$

where (w_s,x_s) and (w_t,y_t) are the weights and points for the G(E)-point Gauss-Legendre integration rules on $[x_E,x_E+h_E]$ and $[y_E,y_E+h_E]$ respectively. Choices for G(E) are specified in Chapter 7.

6.2 Assembling the linear system

Consider Step 1.a (Section 6). For simplicity, suppose

$$a_E(v,w) = (a(x,y)\ v,w)_E\ ,$$

the bilinear form for a weighted least squares problem, since considerations for general $a_E(\cdot,\cdot)$ are almost identical.

Since $\{\tilde{B}_I|_E\}$ is a subset of

$$\{p_i^E(x)p_m^E(y)\ :\ i=1,\ldots,kmax\ ,\ m=1,\ldots,kmax\ \},$$

sparse versions of the tensor-product assembly procedures in Eisenstat and Schultz [18] and Weiser, Eisenstat, and Schultz [31] are appropriate. If

$$\tilde{B}_I = p_i(x)p_m(y),$$

$$\tilde{B}_J = p_j(x)p_n(y),$$

then, depending on whether a(x,y) and f(x,y) are non-separable, separable, or constant, $\tilde{A}^E$ and $\tilde{f}^E$ can be computed using the formulae

$$(a(x,y)\ \tilde{B}_I,\tilde{B}_J)_E \cong$$

$$\sum_t \langle w_t p_m(y_t)p_n(y_t)\rangle \left[\sum_s \langle w_s p_i(x_s)p_j(x_s)\rangle\ a(x_s,y_t) \right],$$

$$(a_x(x)a_y(y)\ \tilde{B}_I,\tilde{B}_J)_E \cong$$

$$\left[\sum_s \langle w_s p_i(x_s)p_j(x_s)\rangle\ a_x(x_s)\right] \left[\sum_t \langle w_t p_m(y_t)p_n(y_t)\rangle\ a_y(y_t)\right],$$

$$(\tilde{B}_I,\tilde{B}_J)_E \cong \left\langle \left(\sum_s w_s p_i(x_s)p_j(x_s)\right) \left(\sum_t w_t p_m(y_t)p_n(y_t)\right) \right\rangle,$$

$$(f(x,y),\tilde{B}_I)_E \cong \sum_t \langle w_t p_m(y_t)\rangle \left[\sum_s \langle w_s p_i(x_s)\rangle\ f(x_s,y_t) \right],$$

$$(f_x(x)f_y(y),\tilde{B}_I)_E \cong$$

$$\left[\sum_s \langle w_s p_i(x_s)\rangle\ f_x(x_s) \right] \left[\sum_t \langle w_t p_m(y_t)\rangle\ f_y(y_t) \right],$$

$$(1,\tilde{B}_I)_E \cong \left\langle \left(\sum_s w_s p_i(x_s)\right) \left(\sum_t w_t p_m(y_t)\right) \right\rangle.$$

The quantities in angle brackets $\langle\rangle$ can be precomputed independent of E. The quantities in square brackets [] should be computed as intermediate sums in each E. Since

$$\tilde{A}^E_{IJ} = \tilde{A}^E_{(i,m)(j,n)} = \tilde{A}^E_{(i,n)(j,m)} = \tilde{A}^E_{(j,m)(i,n)} = \tilde{A}^E_{(j,n)(i,m)},$$

only entries in $\tilde{A}^E$ with $i \leq j$ and $m \leq n$ should be computed explicitly.

Consider Step 1.b (Section 6). For simplicity, we discuss Step 1.b only for element E in Figure 6-1, when k_E is uniform. Other cases are handled analogously.

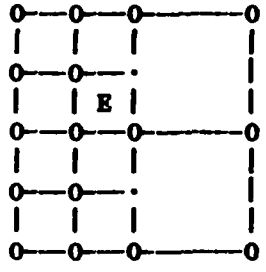


Figure 6-1: An element with one irregular corner

The $\{B_I\}$ nonzero in E have the values

$$p_i^E(x)p_m^E(y) \qquad i=1,\ldots,k,\ i\neq 2\ ,\ m=1,\ldots,k,$$

$$p_2^E(x)p_m^{f(E)}(y) \qquad\qquad\qquad\qquad m=1,\ldots,k.$$

Suppose

$$p_m^{f(E)}(y) = \sum_{L=1}^{\max(2,m)} v_{m,L}\, p_L^E(y). \tag{6.2}$$

Since refinement is regular, the values $\{v_{m,L}\}$ can be precomputed independent of E, and A^E and f^E can be obtained from the formulae

$$A^E_{(i,m)(j,n)} = \begin{cases} \tilde{A}^E_{(i,m)(j,n)} & i\neq 2,\ j\neq 2, \\ \sum_L v_{m,L}\, \tilde{A}^E_{(i,L)(j,n)} & i=2,\ j\neq 2, \\ A^E_{(j,n)(i,m)} & i\neq 2,\ j=2, \\ \sum_P v_{n,P} \left[\sum_L v_{m,L}\, \tilde{A}^E_{(i,L)(j,P)} \right] & i=2,\ j=2, \end{cases}$$

and

$$f^E_{(i,m)} = \begin{cases} \tilde{f}^E_{(i,m)} & i\neq 2, \\ \sum_L v_{m,L}\, \tilde{f}^E_{(i,L)} & i=2. \end{cases}$$

The resulting computation is relatively expensive when k_E is small. When a(x,y) is separable, fewer multiplies are required if Step 1.b is omitted, and A^E and f^E are formed directly.

6.3 Solving the linear system

Consider Step 2 (Section 6), solving the linear system A U = f. In this section, we investigate the complexity of sparse direct elimination with nested-dissection-type ordering of the unknowns for the three classes of problems discussed in discussed in Section 5.7.

6.3.1 A smooth solution

As in Section 5.7, suppose U is smooth everywhere in Ω. Suppose k and h are uniform, and there are $N := h^{-2}$ elements. There are (k-4)(k-3)/2 element basis functions nonzero in each E. In sparse

Gaussian elimination, if the element basis functions are ordered first, their elimination induces no fill-in (i.e., they are statically condensed [28]). This initial elimination requires $\tilde{=} N((k-4)(k-3)/2)^3/6$ multiplies and $\tilde{=} N((k-4)(k-3)/2)^2/2$ storage. Although this cost dominates when $k >> N$, in solving problems to usual accuracy, elimination of the element basis functions is relatively inexpensive. There are $\tilde{=} (2k-3)N$ remaining basis functions (k-2 side basis functions associated with each distinct element side, and 1 vertex basis function associated with each distinct vertex).

Consider a nested dissection ordering of the side and vertex basis functions, in conjunction with sparse elimination of A (e.g., George and Liu [22]). A sample nested dissection ordering is shown in Figure 6-2.

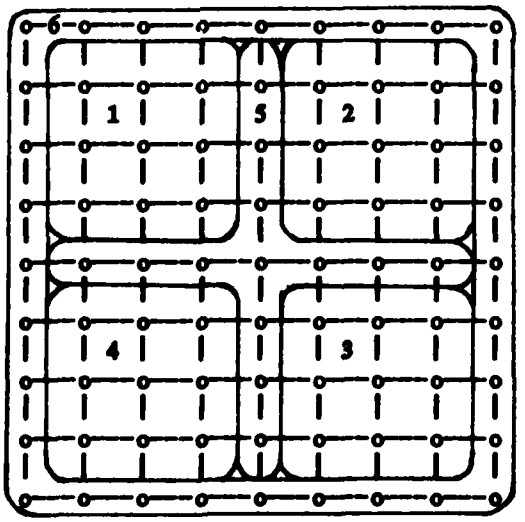


Figure 6-2: Nested dissection ordering with a smooth solution
The unknowns are ordered according to the groups numbered 1-6. Groups

1-4 correspond to basis functions on uniform submeshes with N/4 elements: in each group the unknowns are recursively ordered using nested dissection on the submesh. Since group 5 separates the unknowns in the first four groups, the number of multiplies to eliminate the unknowns in groups 1-5 is, to dominant-order terms, bounded by M(N), where

$$M(N) = 4\ M(N/4) + C\ (kN^{1/2})^3.$$

The rightmost term is a bound on the number of multiplies needed to densely eliminate the unknowns in group 5. The solution to this recurrence relation is $O(k^3N^{3/2})$ (e.g., Rose and Whitten [26], Theorem 1), as is the operation count to eliminate the remaining unknowns in group 6. Thus, the expected operation count, neglecting static condensation cost, is $O(k^3N^{3/2})$. Similarly, the expected storage is $O(k^2N\log(k^2N))$.

6.3.2 A point singularity

As in Section 5.7, suppose u has an isolated singularity at the point $\sigma=(0,0)$. Suppose there are unrefined elements at levels 1 to L in the mesh, and at each level only the elements which touch σ are present. Suppose k is uniform.

There are $N \tilde{=} 3L$ unrefined elements in the mesh. After static condensation, $\tilde{=} (2k-3)N$ side and vertex basis functions remain.

Consider a recursive ordering of the side and vertex basis

functions, in conjunction with sparse elimination of A. A sample ordering is shown in Figure 6-3.

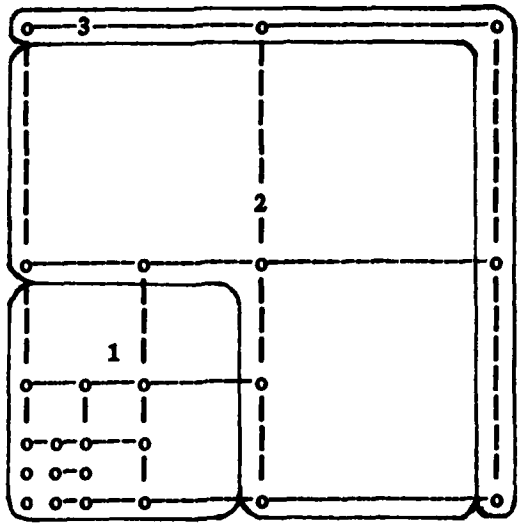


Figure 6-3: Recursive ordering with a point singularity

The unknowns are ordered according to the groups numbered 1-3. Group 1 corresponds to basis functions on a submesh with N-3 elements: the unknowns in group 1 are recursively ordered on the submesh. The number of multiplies to eliminate the unknowns in groups 1 and 2 is, to dominant-order terms, bounded by M(N), where

$$M(N) = M(N-3) + C\,k^3.$$

The rightmost term is a bound on the number of multiplies needed to densely eliminate the unknowns in group 2. The solution to this recurrence relation is $O(k^3N)$, and the operation count to eliminate the remaining unknowns in group 3 is $O(k^3)$. Thus, the expected operation

count, neglecting static condensation cost, is $O(k^3N)$. Similarly, the expected storage is $O(k^2N)$.

6.3.3 A line singularity

As in Section 5.7, suppose u has a line singularity along the line $\sigma = \{(0,y)\}$. Suppose there are unrefined elements at levels 1 to L in the mesh, where at each level the elements which touch the line σ are present. Suppose k is uniform.

There are $N \tilde{=} 3(2^L)$ unrefined elements in the mesh. After static condensation, $\tilde{=} (2k-3)N$ side and vertex basis functions remain.

Consider a nested-dissection-type ordering of the side and vertex basis functions, in conjunction with sparse elimination of A. A sample ordering is shown in Figure 6-4. The unknowns are ordered according to the groups numbered 1-4. Groups 1 and 2 correspond to basis functions on submeshes with (N-2)/2 elements: in each group the unknowns are recursively ordered on the submesh. Since group 3 separates the unknowns in the first two groups, the number of multiplies to eliminate the unknowns in groups 1-3 is, to dominant-order terms, bounded by M(N), where

$$M(N) = 2\,M(N/2) + C\,(k\log N)^3.$$

The rightmost term is a bound on the number of multiplies needed to densely eliminate the unknowns in group 3. The solution to this recurrence relation is $O(k^3N)$ (e.g., Rose and Whitten [26]), and the

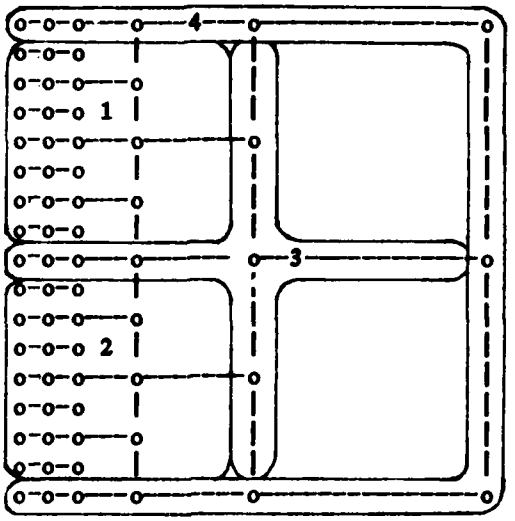


Figure 6-4: Nested-dissection-type ordering with a line singularity

operation count to eliminate the remaining unknowns in group 4 is $O((k\log N)^3)$. Thus, the expected operation count, neglecting static condensation cost, is $O(k^3N)$. Similarly, the expected storage is $O(k^2N)$.

6.3.4 Discussion

The operation counts and storage for the problems with singularities are optimal-order in the number of elements N. We expect similar operation counts for more general problems with point or line singularities. The optimality for the line singularity problem depends on the fact that we can systematically find sets of $O(N^{1/3-\varepsilon})$ elements ($\varepsilon > 0$) which separate the remaining elements into two or more disjoint sets of equal size.

We do not know an efficient (i.e., O(N)), automatic, way of finding orderings such as those presented. See George and Liu [22], Sections 7.2 and 8.2, for work in this direction.

In the example problems, the orderings of the interior unknowns correspond to the reverse of the order in which the edges associated with the unknowns are created. This correspondence does not always hold. (For example, suppose u has a point singularity at (1/2,0).)

6.4 Computing the error indicators

Consider Step 3 (Section 6), the computation of the error indicators $|||\tilde{\varepsilon}|||_E$, $|||\underline{\varepsilon}|||_E$, $|||\bar{\varepsilon}|||_E$, or $\underline{\underline{\varepsilon}}_E$. Recall

$$\tilde{F}_E(v) := (f,v)_E + \langle [a\partial U/\partial n], v\rangle_{\partial E} - a_E(U,v).$$

First, suppose E has no irregular corners. Then the possible quantities to be computed in Step 3 are

1. $\{a_E(\tilde{B}_I,\tilde{B}_J)\}$, where $\{\tilde{B}_I\}$ is the basis for $\bar{S}_E$ which would result if E had no irregular corners. These quantities can be computed as in Step 1.a.

2. $\{(f,\tilde{B}_I)_E : \tilde{B}_I \text{ in } \bar{S}_E - S_E\}$. These quantities can be computed as in Step 1.a.

3. $\{a_E(U,\tilde{B}_I) : \tilde{B}_I \text{ in } \bar{S}_E - S_E\}$. First, convert the representation

$$U(x,y)|_E = \sum_J U_J B_J(x,y)$$

into the form

$$U(x,y)|_E = \sum_J \tilde{U}_J \tilde{B}_J(x,y).$$

For example, for element E in Figure 6-1, when k is uniform, if

$$U(x,y)|_E = \sum_{i\neq 2} \sum_m U_{(i,m)}\ p_i^E(x)\ p_m^E(y) + \sum_m U_{(2,m)}\ p_2^E(x)\ p_m^{f(E)}(y),$$

then

$$\tilde{U}_{(i,m)} = \begin{cases} U_{(i,m)} & , i \neq 2, \\ \sum_{L=\underline{L}(m)}^{k} v_{L,m}\ U_{(i,L)} & , i = 2, \end{cases}$$

where

$$\underline{L}(m) = \begin{cases} m & , m \neq 2, \\ 1 & , m = 2. \end{cases}$$

Other cases are handled analogously.

Then, for $|||\tilde{\varepsilon}|||_E$, compute $\{a_E(U,\tilde{B}_I)\}$ with a matrix-vector multiply. For $|||\underline{\varepsilon}|||_E$ or $|||\bar{\varepsilon}|||_E$, compute $\{U(x_s,y_t)\}$ by the formula

$$U(x_s,y_t) = \sum_m p_m(y_t)\ [\ \sum_i p_i(x_s)\ \tilde{U}_{(i,m)}\ ], \qquad (6.3)$$

and then use the formula

$$(a(x,y)\ U,\tilde{B}_J)_E \tilde{=} \sum_t \langle w_t p_n(y_t)\rangle \left[\sum_s \langle w_s p_i(x_s)\rangle \left[U(x_s,y_t)a(x_s,y_t) \right] \right].$$

4. $\{\langle [a\partial U/\partial n], \tilde{B}_I\rangle_{\partial E}\}$. In a preprocessing step, compute and store the coefficients of the one-dimensional polynomials $\partial U/\partial n|_s$ on each element side. Then, for example, on the left side of element E, compute $\{a\partial U/\partial n(x_E^+, y_t)|_E\}$ and $\{a\partial U/\partial n(x_E^-, y_t)|_{E'}\}$, where element E', if it exists, shares part of the left side of E. Form

$$\langle [a\partial U/\partial n], \tilde{B}_I\rangle_{\partial E(\text{left})} \tilde{=}$$

$$\sum_t \langle w_t p_m(y_t)/2\rangle [a\partial U/\partial n(x_E^+, y_t)|_E - a\partial U/\partial n(x_E^-, y_t)|_{E'}]$$

for each $\tilde{B}_I = p_l(x)p_m(y)$. Other cases are handled analogously.*

5. $|||\tilde{\varepsilon}|||_E^2$, $|||\underline{\varepsilon}|||_E^2$, or $|||\bar{\varepsilon}|||_E^2$. Suppose the system for

---

*The storage required can be reduced by intermixing the preprocessing step with error indicator computation.

the coefficients of $\tilde{\varepsilon}_E$ is $A\,\tilde{v} = f$. Compute a lower triangular matrix L such that $L\,L^T = A$, and solve $L\,y = f$. (For $\underline{\varepsilon}_E$ and $\bar{\varepsilon}_E$, L is independent of h_E.) Then compute $y^T y = |||\tilde{\varepsilon}|||_E^2$.

6. $||f-LU||_E^2$, for $\underline{\varepsilon}_E$. First compute $\{U(x_s,y_t)\}$ as in (6.3), and then compute

$$||f-LU||_E^2 \tilde{=} \sum_{s,t} (a(x_s,y_t)U(x_s,y_t) - f(x_s,y_t))^2.$$

7. $|[a\partial U/\partial n]-a\partial U/\partial n|_{\partial E}^2$, for $\underline{\varepsilon}_E$. Preprocess $\partial U/\partial n|_s$ as in the computation of $\{\langle[a\partial U/\partial n],\tilde{B}_I\rangle_{\partial E}\}$. Then, for example, on the left side of E, compute

$$|[a\partial U/\partial n]-a\partial U/\partial n|_{\partial E(\text{left})}^2 \tilde{=}$$

$$\sum_t \langle w_t/4\rangle \,(a\partial U/\partial n(x_E^+,y_t)|_{\partial E} + a\partial U/\partial n(x_E^-,y_t)|_{\partial E'})^2.$$

Now suppose E has an irregular corner. For simplicity, we only consider element E in Figure 6-1, when k and $\bar{k} > k$ are uniform. Other cases are handled analogously. Suppose

$$p_m^E(y) = \sum_{P=1}^{\max(2,m)} w_{m,P}\, p_P^{f(E)}(y).$$

Since refinement is regular, the values $\{w_{m,P}\}$ can be precomputed independent of E, and the right hand side components have values

$$f_{(i,m)} = \tilde{F}_E(\tilde{B}_{(i,m)})$$

when $\tilde{B}_{(i,m)}$ is in $\bar{S}_E - S_E$ and $i \neq 2$, and

$$f_{(i,m)} = \sum_{P=k+1}^{m} w_{m,P} \left[\sum_{L=1}^{P} v_{P,L} \tilde{F}_E(\tilde{B}_{(i,L)}) \right]$$

when $k < m \leq \bar{k}$ and $i = 2$. The rest of the computations are unchanged.

CHAPTER 7

Numerical results

7.1 Introduction

In this chapter, we present numerical results obtained using prototype codes written in FORTRAN, which implement some of the methods discussed in Chapters 2 - 5. The problem set, containing one problem from each of the three classes considered in Section 5.7, is specified in Section 7.2. In Section 7.3, we present the error behavior, which is in fair accord with Section 5.7. In Section 7.4, we present the error estimator behavior: the error estimators are usually accurate to within a factor of two. In some cases, some of the estimators appear to converge to the norm of the true error. In Section 7.5, we present more numerical evidence of error estimator convergence.

7.2 The problem set

Prototype codes were written in FORTRAN, implementing some of the methods discussed in Chapters 2 - 5. Numerical tests were conducted to investigate the behavior of the methods.

The test problem set consisted of three Poisson problems with

Dirichlet boundary conditions,

$$-D_x^2 u - D_y^2 u = f(x,y) \quad \text{in the interior of } \Omega,$$

$$u(x,y) = g(x,y) \quad \text{on } \partial\Omega,$$

with solutions specified in Table 7-1.

Problem	Solution class	Solution
1	smooth	$u(x,y) = \cos(e^x + xy)$
2	point singularity	$u(x,y) = r^{1/2} \sin(\theta/2)$, $x = r\cos(\theta)$, $y = r\sin(\theta)$
3	line singularity	$u(x,y) = \begin{cases} 0 & t \leq -1/2 \\ \sin(\pi t) & -1/2 < t < 1/2 \\ 1 & 1/2 \leq t \end{cases}$, $t = 5x - y - 1$

Table 7-1: Test problems

The solution to Problem 1 is in $H^k(\Omega)$ for all k. The solution to Problem 2 is the fundamental eigenfunction for a crack problem with interior angle 2π; this function is in $H^{3/2-\varepsilon}(\Omega)$ for any $\varepsilon > 0$, but not in $H^{3/2}(\Omega)$. The solution to Problem 3 is a smeared shock wave with shock-width 1/5 in x; this function is in $H^2(\Omega)$, but not in $H^{2+\varepsilon}(\Omega)$ for

any $\varepsilon > 0$.

We tested the seven refinement methods specified in Table 7-2. All adaptive sequences of finite element spaces started out with 2X2 meshes. The adaptive sequences for methods + and X started out with $k_E = 2$.

Method	h_E	k_E
U	uniform (1/2,...,1/32)	= 2
K	= 1/2	adaptive (Section 5.4)
2	adaptive (Section 5.3)	= 2
3	adaptive (Section 5.3)	= 3
4	adaptive (Section 5.3)	= 4
+	adaptive	adaptive (Section 5.6)
X	adaptive	adaptive (Section 5.4)

Table 7-2: Refinement methods

We used three heuristic refinement procedures, based on the formulae in Chapter 5. In the "true error" procedure, we set $\xi_E = |||e|||_E^2$. In the "approximate error" procedure, we set $\xi_E = |||\bar{e}|||_E^2$, with $\bar{S}$ composed of piecewise polynomials one order higher than S. In the "residuals and jumps" procedure, we set $\xi_E = \xi_E^2$. For simplicity, we set $\lambda_{CUT}^{EST} = \lambda_1/2$. To avoid ill-conditioned linear systems, we restricted $k_E \leq 7$. Our tests were run in single precision on a DEC-System 2060 computer.

G(E)-point Gauss-Legendre quadrature rules (6.1) were used to approximate integrals in each element E. In the original assembly phase, $G(E) = kmax_E$, where

$$kmax_E := \max\{k_{E'} : E' \text{ is in } N(E)\}.$$

In the error indicator phase, $G(E) = kmax_E+1$. In the evaluation of $|||e|||_E$, $G(E) = kmax_E+3$. With these values, for problems with smooth solutions, the errors due to numerical quadrature are asymptotically negligible (e.g., Ciarlet [17], Section 4.1).

7.3 Error behavior

In this section, the resulting errors for the adaptive sequences of finite element spaces are depicted on $\log_{10}$-$\log_{10}$ scales. The x-coordinate of each data point is the number of nonzeroes in the stiffness matrix*, and the y-coordinate is $|||e|||$. Since the true work generally grows faster in k than the number of nonzeroes, the measure of work is biased in favor of high-order methods.

Figure 7-1 presents the errors due to applying the "true error"

---

*Our prototype codes should be made more efficient before computer time is used as the measure of work.

procedure to Problem 1. Since u(x,y) is smooth, nonuniform mesh refinement does not improve the asymptotic convergence rate. The slopes of the lines correspond to $O(h^{k-1})$ error and $O(h^{-2}k^4)$ work. Since k increases for methods K, +, and X, these methods achieve faster than polynomial convergence. Methods + and X perform some initial mesh rfinement (see Figure 7-2, with k_E labelled in each element), so they are somewhat less efficient than method K.

Figures 7-3 and 7-4 present the errors due to applying the "approximate error" and "residuals and jumps" procedures to Problem 1. The errors are similar to the ones shown in Figure 7-1.

Figure 7-5 presents the errors due to applying the "true error" procedure to Problem 2. The slope of the line for method U corresponds to an $O(h^{1/2})$ convergence rate and an $O(h^{-2})$ work count. Method K has almost the same line, with slope corresponding to an $O(k^{-1})$ convergence rate and an $O(k^4)$ work count.* The slopes of the lines for methods 2, 3,

---

*With the work count proportional to the number of unknowns, method K would appear twice as efficient as method U (see [11]). With a more realistic work count, method K would asymptotically be the least efficient method considered.

and 4 are the same as the slopes for the same methods in Figure 7-1. Local mesh refinement, in effect, transforms variables so the problem looks like it has a smooth solution on a uniform mesh. Methods + and X are fairly efficient in all cases, refining the mesh near the singularity, and refining the order where the solution is smooth (see Figure 7-6).

Figures 7-7 and 7-8 present the errors due to applying the "approximate error" and "residuals and jumps" procedures to Problem 2. The errors are similar to the ones shown in Figure 7-5.

Figure 7-9 presents the errors due to applying the "true error" procedure to Problem 3. The slope of the line for method U corresponds to an $O(h)$ convergence rate and an $O(h^{-2})$ work count. The line for method K appears to level out, reflecting an $O(k^4)$ work count and slow convergence in k. The slope of the line for method 2 reflects an improving ($< N^{-1/2}$) convergence rate and an $O(N)$ work count, where N is the number of elements. At the modest accuracies required, the mesh is refined not only along the line singularities at the edges of the boundary layer, but in the boundary layer as well (see Figure 7-10). We expect the slope of the line for method 2 to approach -1 asymptotically, reflecting an $O(N^{-1})$ convergence rate. The slopes of the lines for methods 3, 4, +, and X reflect convergence rates tending toward $O(N^{-1})$. Figure 7-11 depicts a sample grid for method +. Note that the shape of the boundary layer cannot be discerned from the grid.

Figures 7-12 and 7-13 present the errors due to applying the "approximate error" and "residuals and jumps" procedures to Problem 3. The errors are similar to the ones shown in Figure 7-9.

In this test, methods + and X appear to be the most robust, since they are fairly efficient for all three problems, regardless of whether the errors are large or small. Each of the other methods does poorly for some problems in some cases. Table 7-3 summarizes the observed efficiency α, where

$$|||e||| \tilde{=} (\text{the number of nonzeroes})^{-\alpha}.$$

Method	Problem 1	Problem 2	Problem 3
U	1/2	1/4	$< 1/2$
K	(k-1)/2	1/4	$\tilde{=} 1/2$
2	1/2	1/2	$> 1/2$
3	1	1	$\tilde{=} 1$
4	3/2	3/2	$\tilde{=} 1$
+	(k-1)/2	(k-1)/2	$\tilde{=} 1$
X	(k-1)/2	(k-1)/2	$\tilde{=} 1$

Table 7-3: Efficiency summary

7.4 Error estimator behavior

Figures 7-14 - 7-19 depict the relative errors $\bar{\rho}$ and $\underline{\rho}$ in the error estimators for the tests run using the "approximate error" and

"residuals and jumps" procedures respectively.

In the "approximate error" procedure, the error is usually estimated within a factor of two ($-.5 \leq \bar{\rho} \leq 1$), and always

$$-.7 \leq \bar{\rho} \leq 1.6 \ .$$

$\bar{\rho}$ appears to converge to zero in methods U and 2 for Problem 1, and in methods 2 and 3 for Problem 2 (see Section 7.5). For Problem 1, $\bar{\rho}$ varies somewhat erratically in methods + and X, especially when k_E is large. This may be partly due to ill-conditioning.

In the "residuals and jumps" procedure, the error is always estimated within a factor of two. In particular, in method 3,

$$-.25 \leq \underline{\rho} \leq .20 \ ,$$

and $\underline{\rho}$ appears to converge to zero. In methods + and X, $\underline{\rho}$ varies more smoothly than $\bar{\rho}$. Thus, in this test, the "residuals and jumps" procedure appears to be slightly more robust than the "approximate error" procedure.

7.5 Error estimator convergence

As further evidence of apparent error estimator convergence, we now present the behavior of the four error estimators $|||\tilde{\varepsilon}|||$, $|||\underline{\varepsilon}|||$, $|||\bar{\varepsilon}|||$, and $\underline{\varepsilon}$, with relative errors $\tilde{\rho}$, $\underline{\rho}$, $\bar{\rho}$, and $\underline{\rho}$ respectively, for a sample variable-coefficient problem with a smooth solution. k and $h = 1/n$ are uniform. $\bar{S}$ contains piecewise polynomials one order higher

than S. We have observed similar error estimator behavior for other problems with smooth solutions.

Table 7-4 depicts the relative errors in the error estimators for the Dirichlet problem

$$- D_x(e^{xy}D_x u) - D_y(e^{xy}D_y u) + u/(1+x+y) = f(x,y)$$

in the interior of Ω,

$$u(x,y) = 0 \quad \text{on } \partial\Omega,$$

where $f(x,y)$ is chosen so that $u(x,y) = e^{xy} \sin(\pi x) \sin(\pi y)$.

In Table 7-4, the error estimators are always within a factor of two of $|||e|||$. None of the error estimators converge to $|||e|||$ with k increasing and h fixed, but some appear to converge with h decreasing and k fixed. $|||\tilde{\varepsilon}|||$ and $|||\underline{\varepsilon}|||$ appear to converge when $k \leq 4$. $|||\bar{\varepsilon}|||$ appears to converge when $k < 4$, but does not converge when $k = 4$. The convergence rates are not clear. When $k = 2$ or 3, $\underline{\varepsilon}$ behaves like a constant times $|||e|||$.

	n	$\|\|\|e\|\|\|$	$\tilde{\rho}$	ρ	$\bar{\rho}$	$\underline{\rho}$
k=2	2	.181(+1)	.053	.026	.026	.26
	4	.877	.028	.024	.022	.37
	8	.436	.012	.011	.010	.40
	16	.218	.0039	.0037	.0033	.40
	32	.109	.0011	.0010	.0009	.40
k=3	2	.515	-.46	-.45	-.45	-.15
	4	.857(-1)	-.12	-.11	-.12	-.0088
	8	.194(-1)	-.022	-.022	-.022	.018
	16	.477(-2)	-.0041	-.0038	-.0043	.013
k=4	2	.443	.33	.29	.29	-.44
	4	.426(-1)	.12	.10	.039	-.39
	8	.446(-2)	.043	.038	-.11	-.35
	16	.508(-3)	.019	.018	-.25	-.32
k=5	2	.671(-1)	.35	.34	.34	-.36
	4	.426(-2)	.23	.22	.18	-.42
	8	.256(-3)	.20	.19	-.026	-.48
k=6	2	.118(-1)	.31	.29	.29	-.25
	4	.508(-3)	.23	.22	.099	-.29
	8	.156(-4)	.32	.32	.077	-.26
k=7	2	.213(-2)	.36	.35	.35	-.14
	4	.380(-4)	.36	.35	.31	-.19

Table 7-4: Error estimator behavior

Figure 7-1: Problem 1 - true error

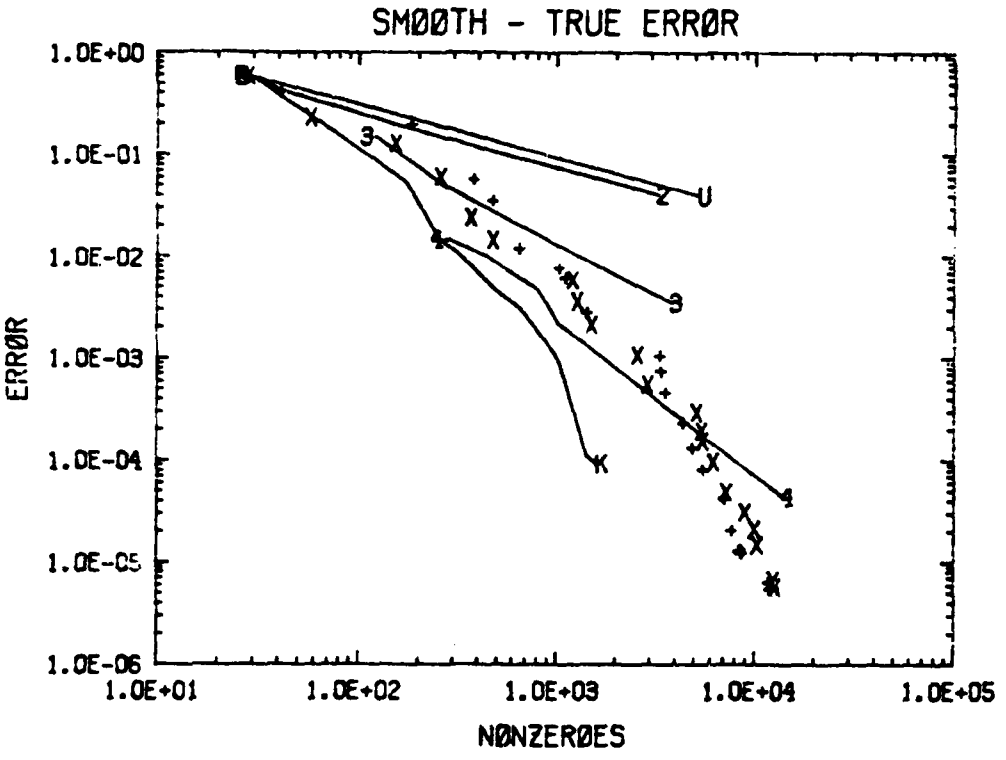


Figure 7-2: Sample grid for problem 1 (method +, residuals and jumps)

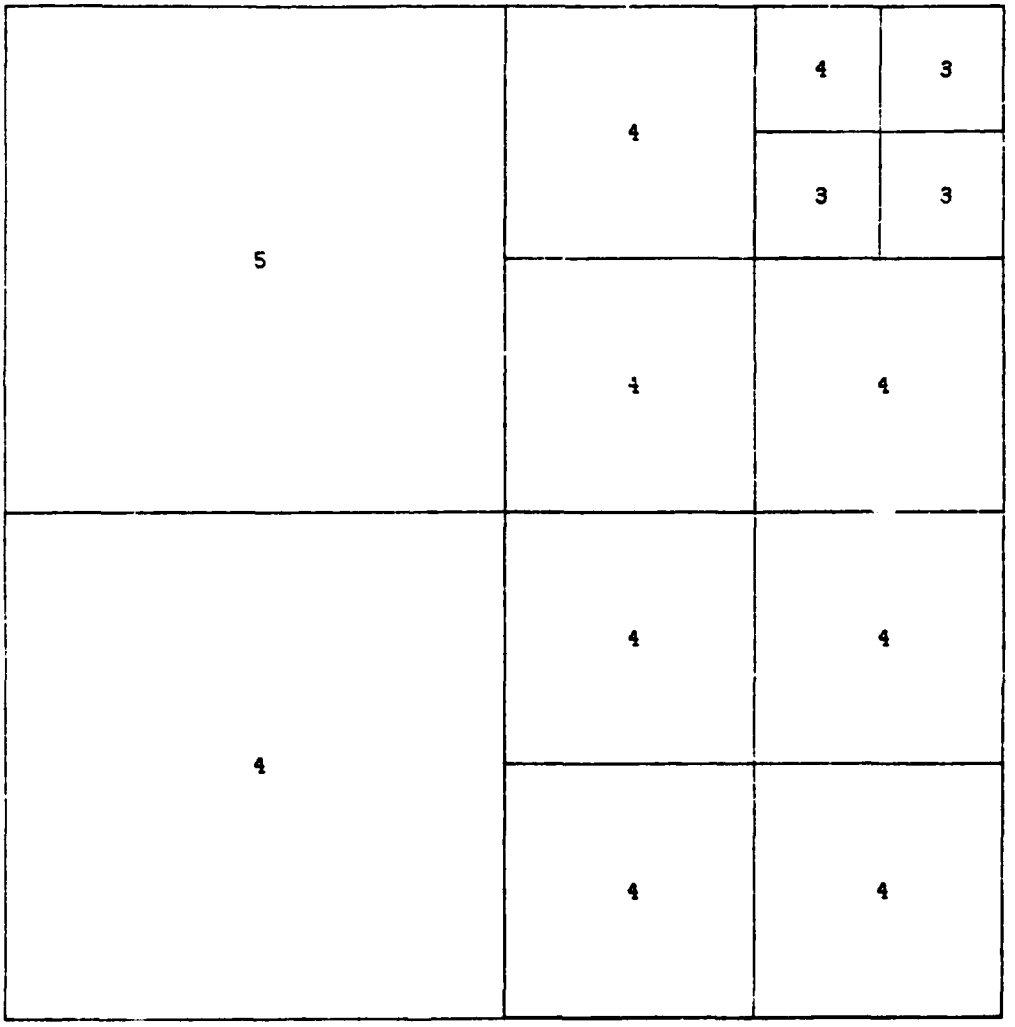


Figure 7-3: Problem 1 - approximate error

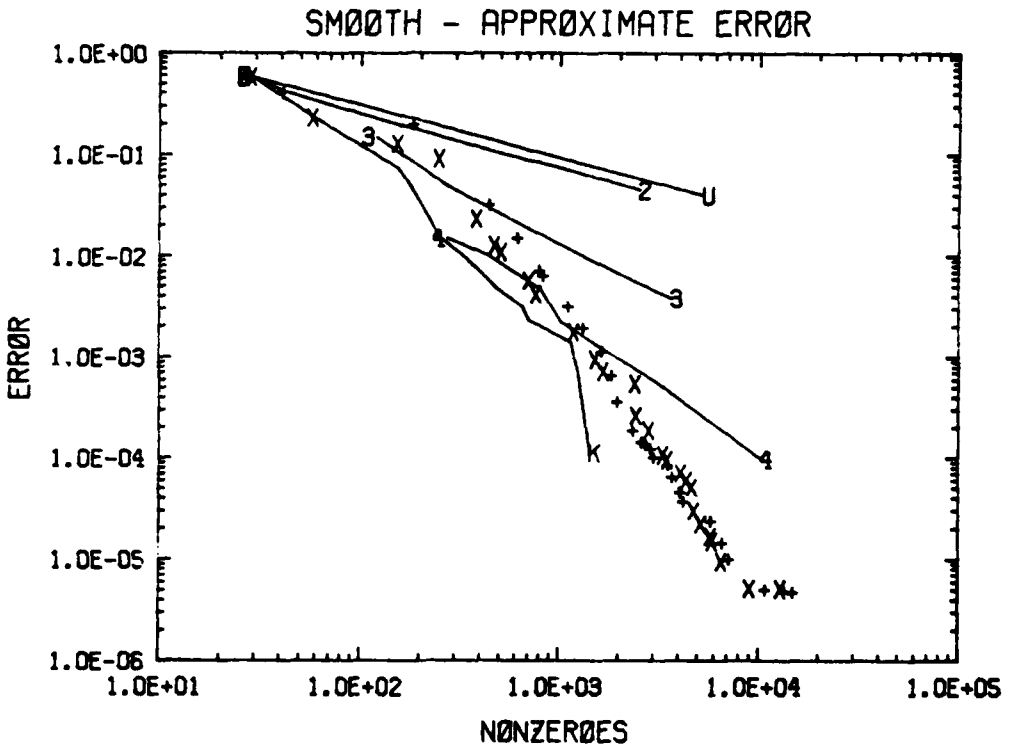


Figure 7-4: Problem 1 - residuals and jumps

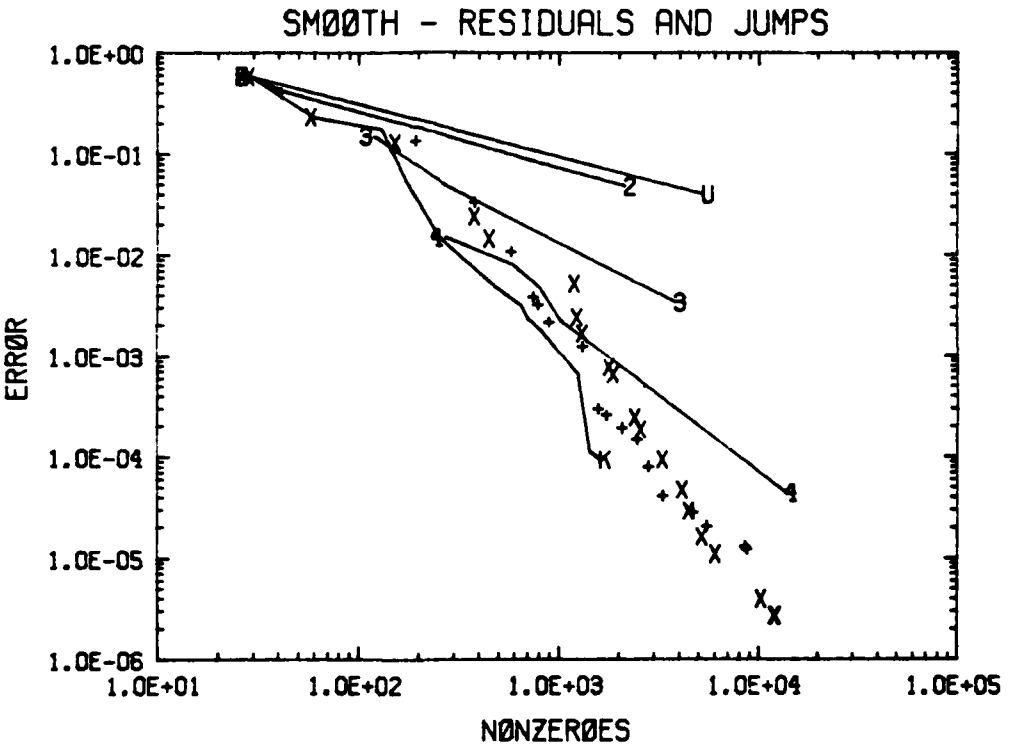


Figure 7-5: Problem 2 - true error

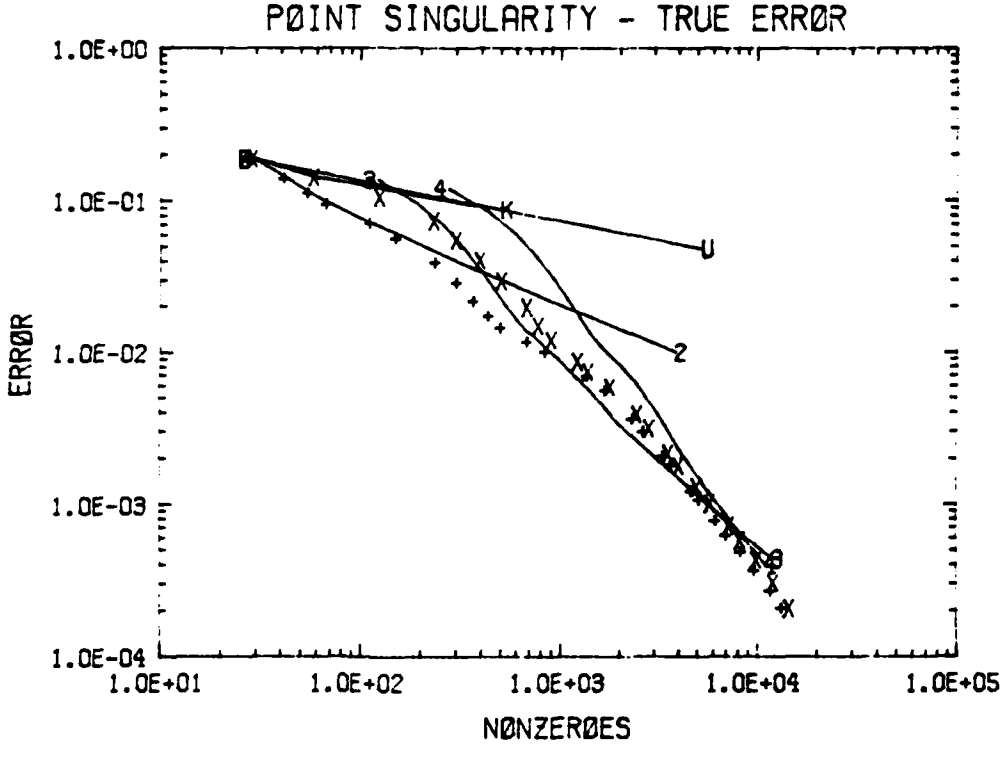


Figure 7-6: Sample grid for problem 2 (method +, residuals and jumps)

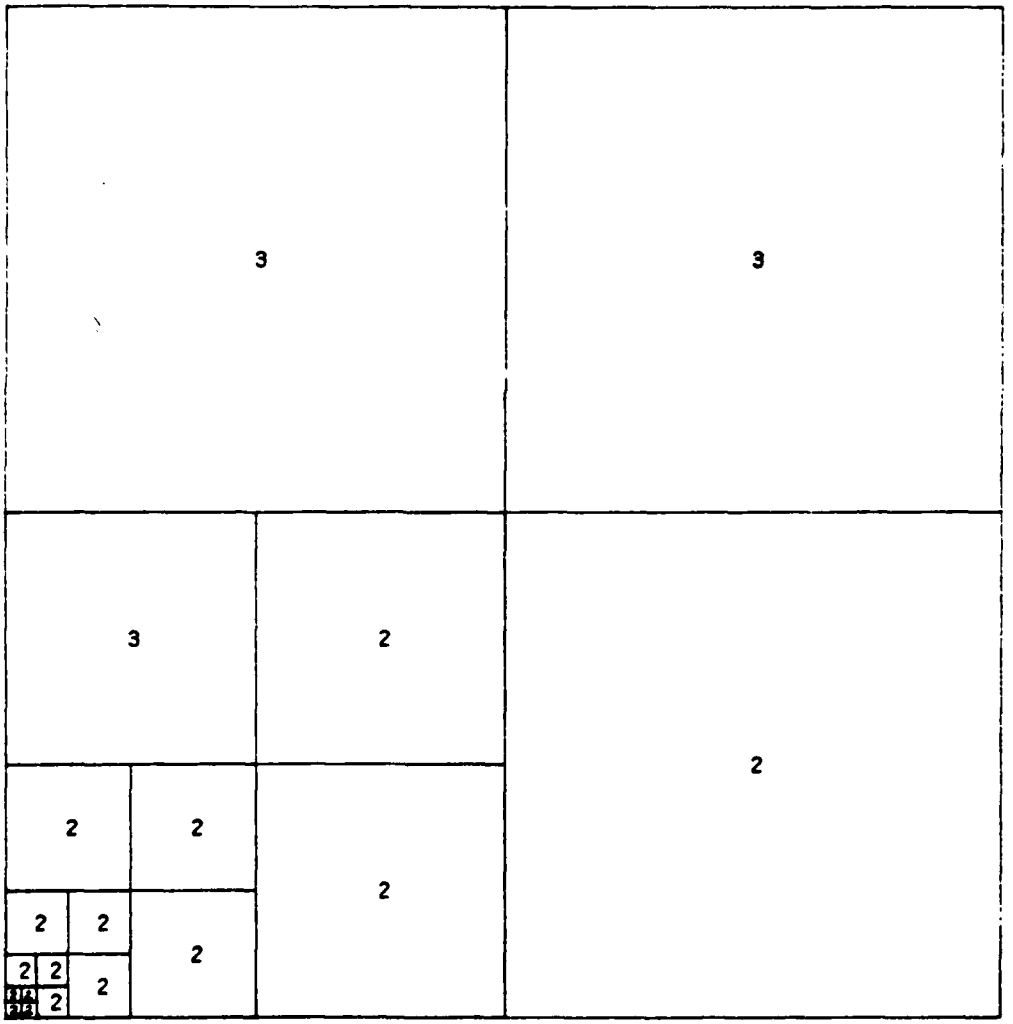


Figure 7-7: Problem 2 - approximate error

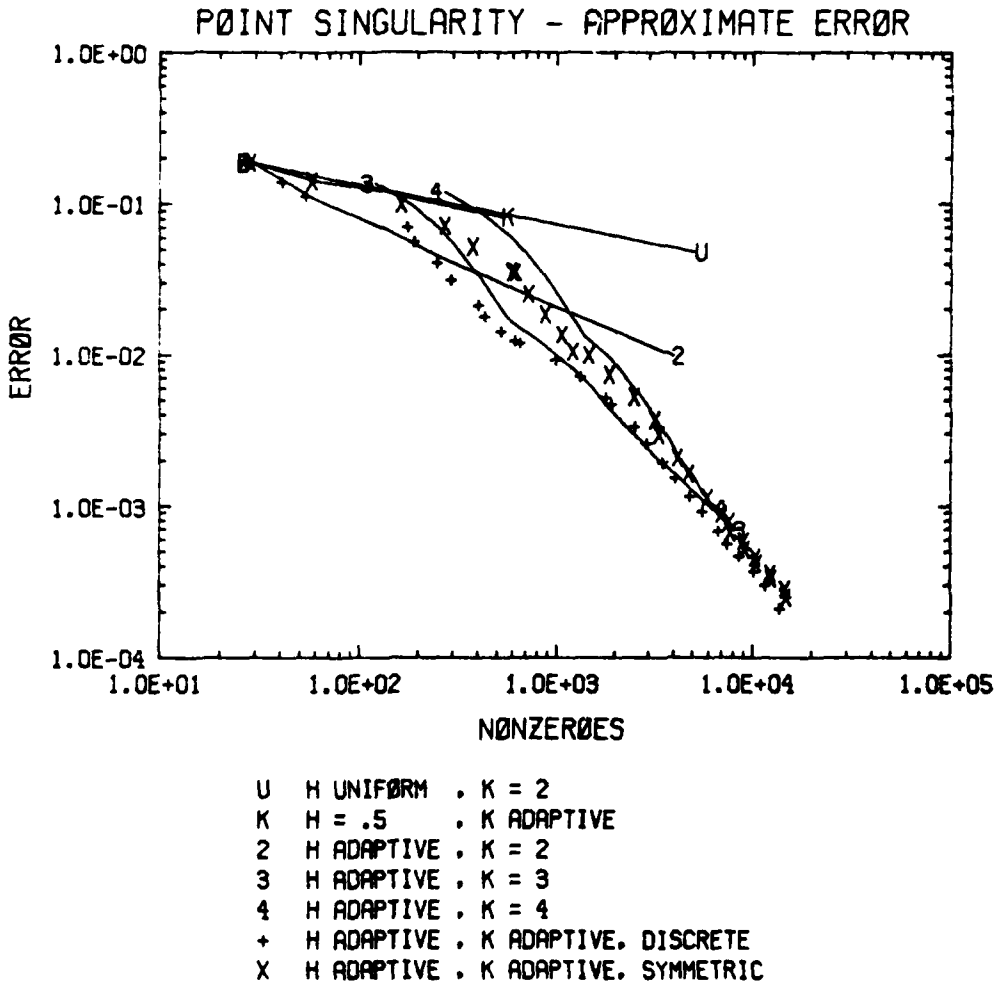


Figure 7-8: Problem 2 - residuals and jumps

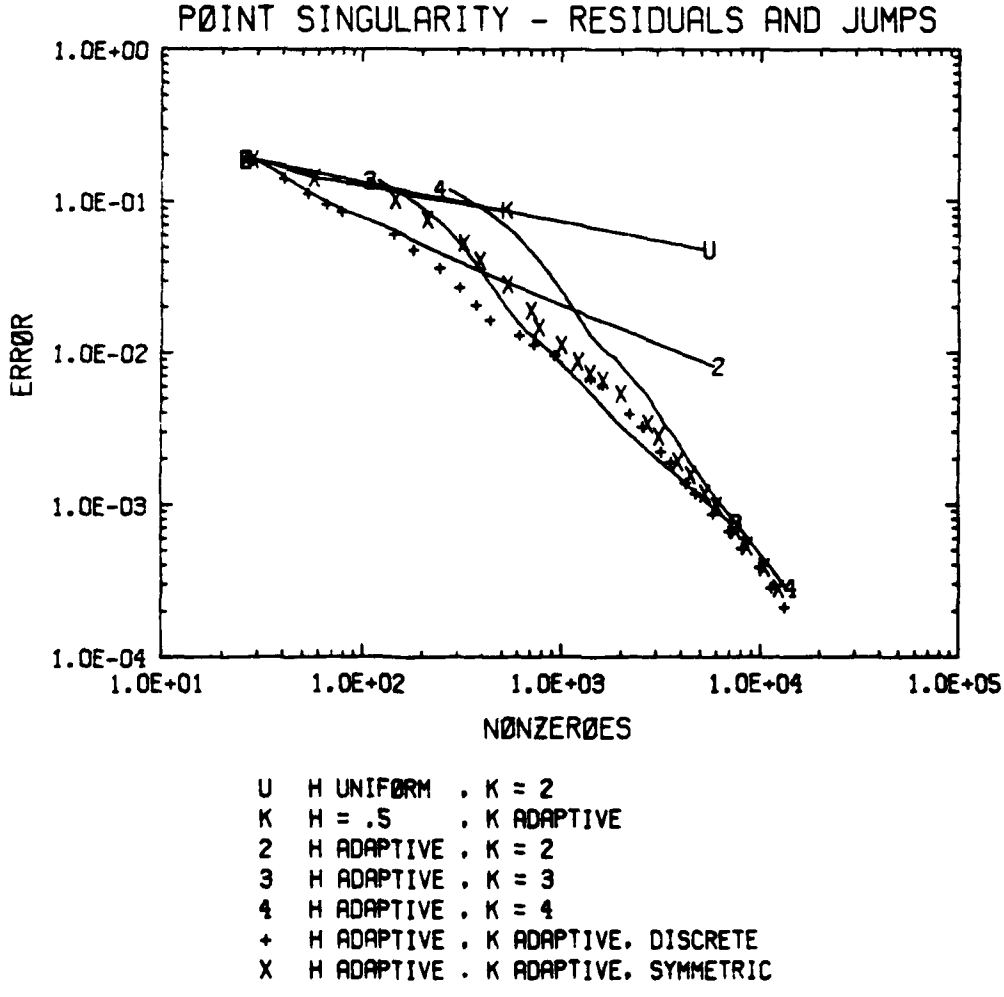


Figure 7-9: Problem 3 - true error

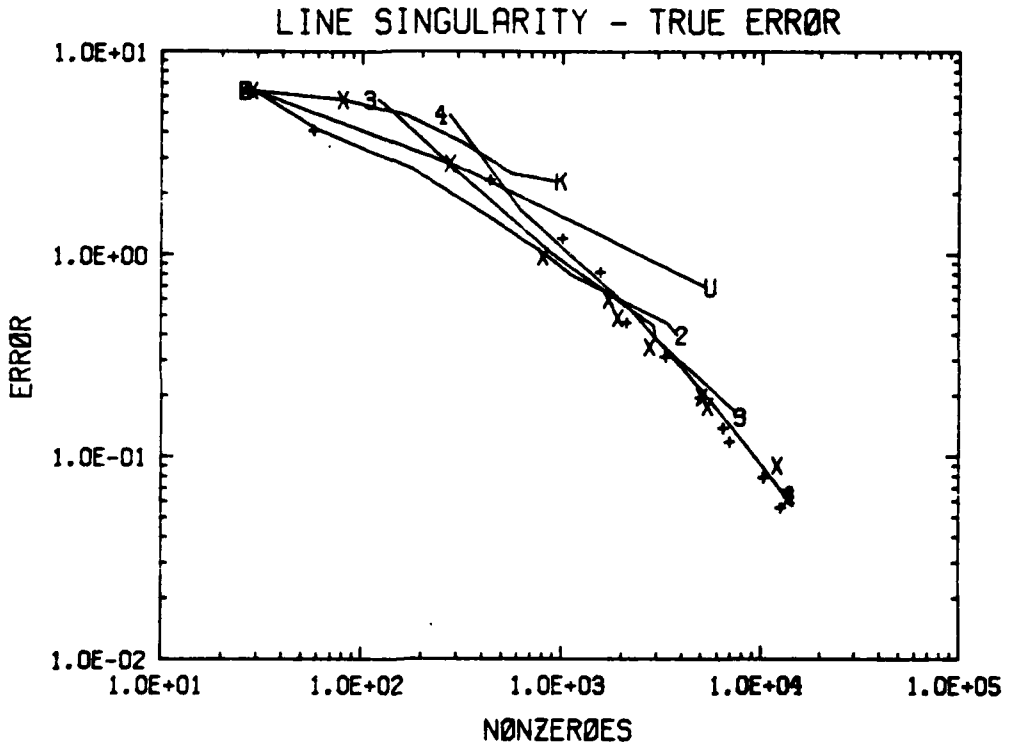


Figure 7-10: Sample grid for problem 3 (method 2, residuals and jumps)

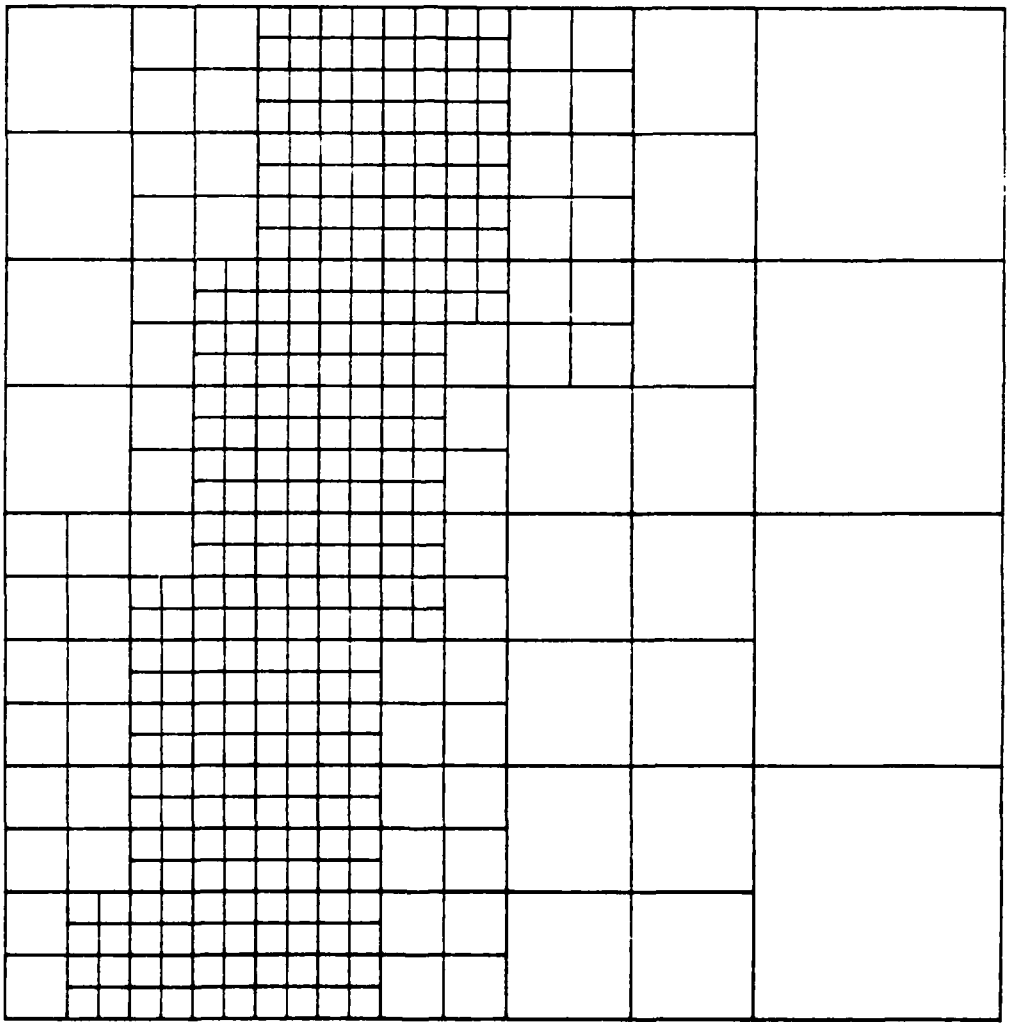


Figure 7-11: Sample grid for problem 3 (method +, residuals and jumps)

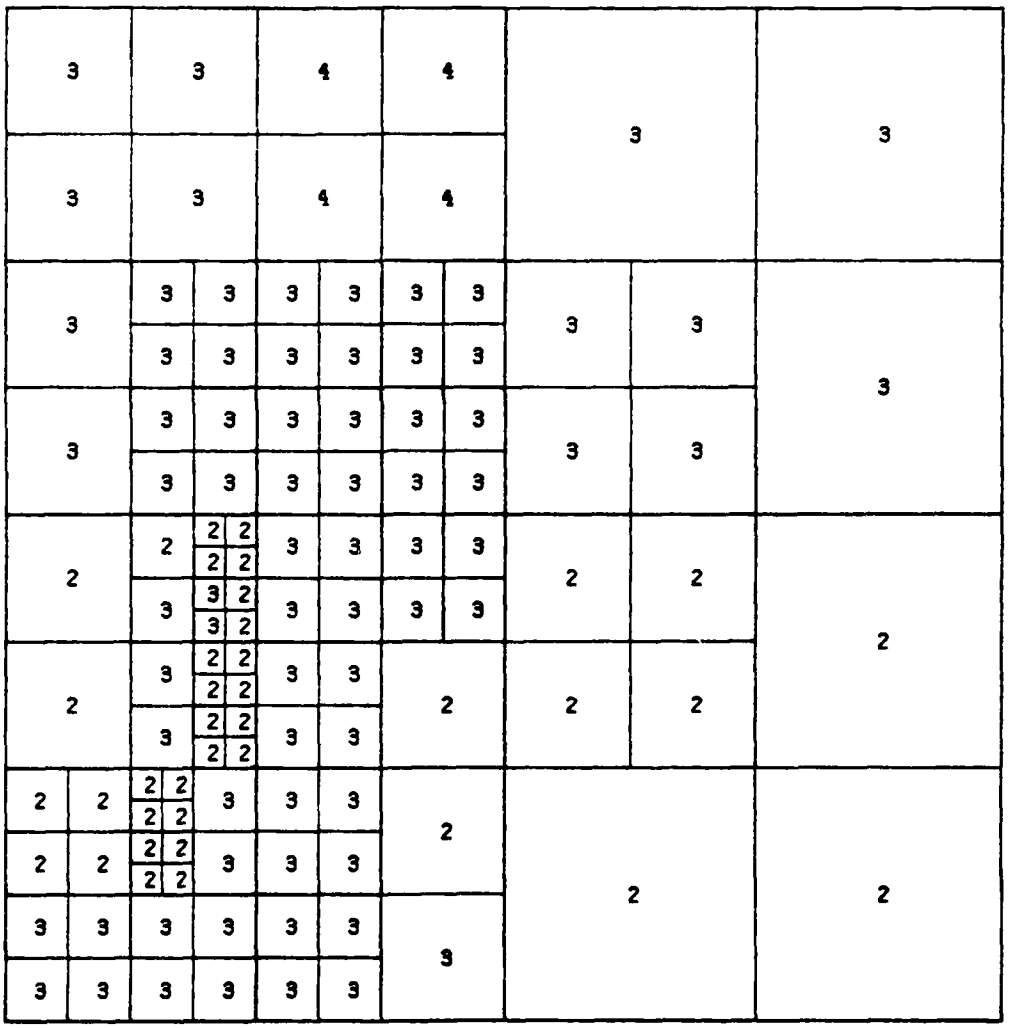


Figure 7-12: Problem 3 - approximate error

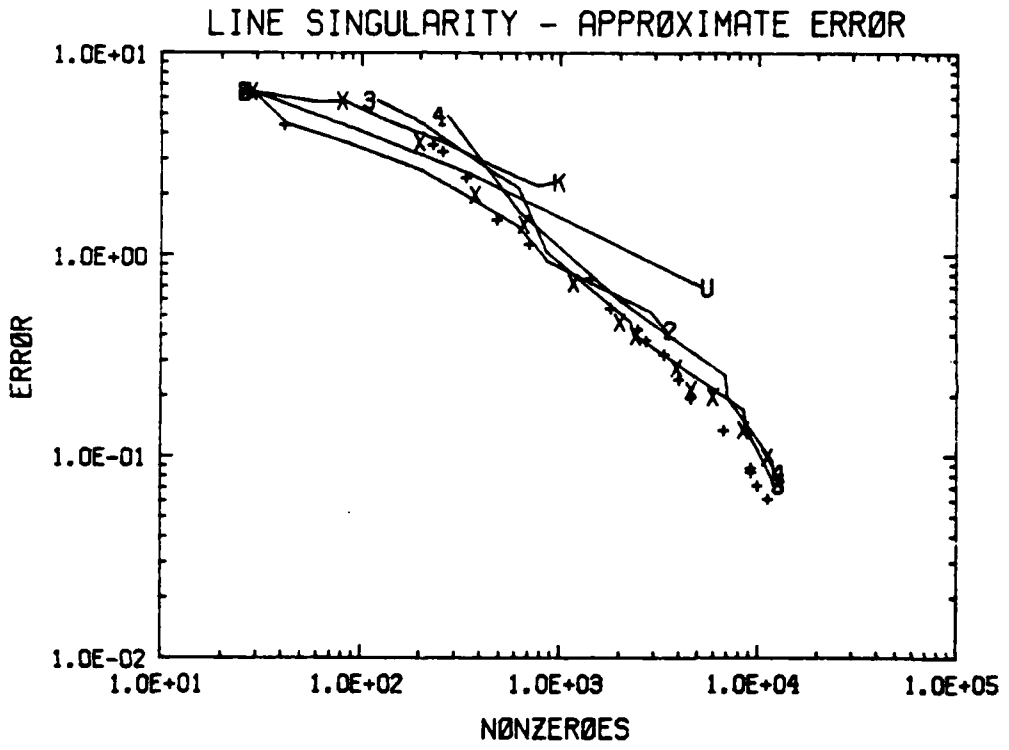


Figure 7-13: Problem 3 - residuals and jumps

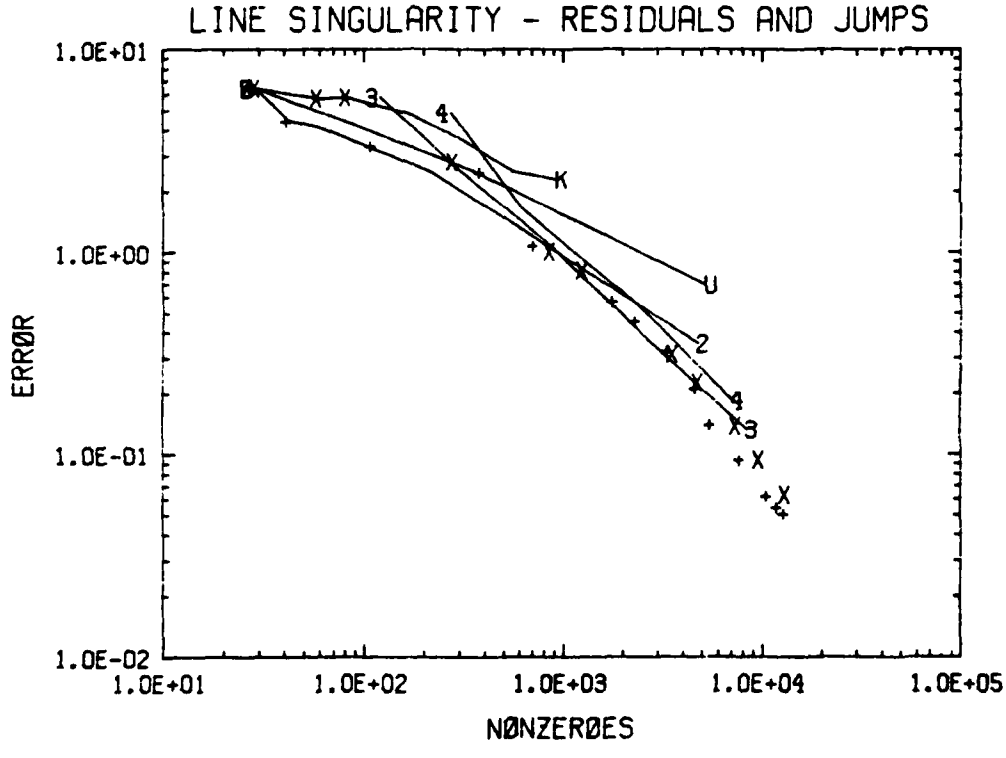


Figure 7-14: Problem 1 - approximate error

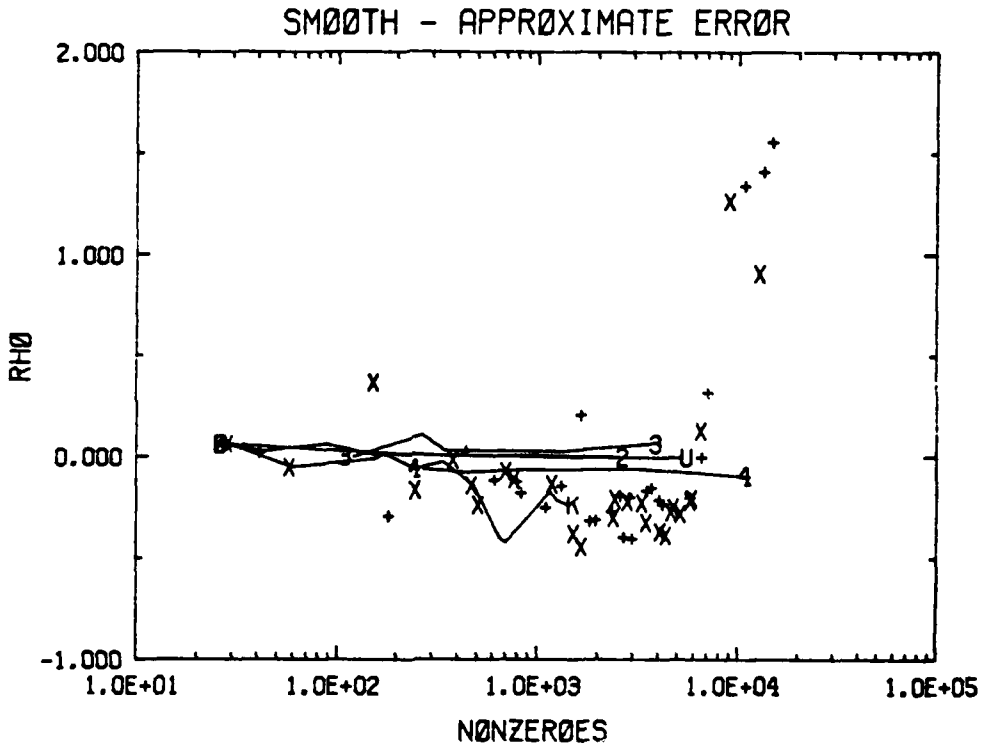


Figure 7-15: Problem 1 - residuals and jumps

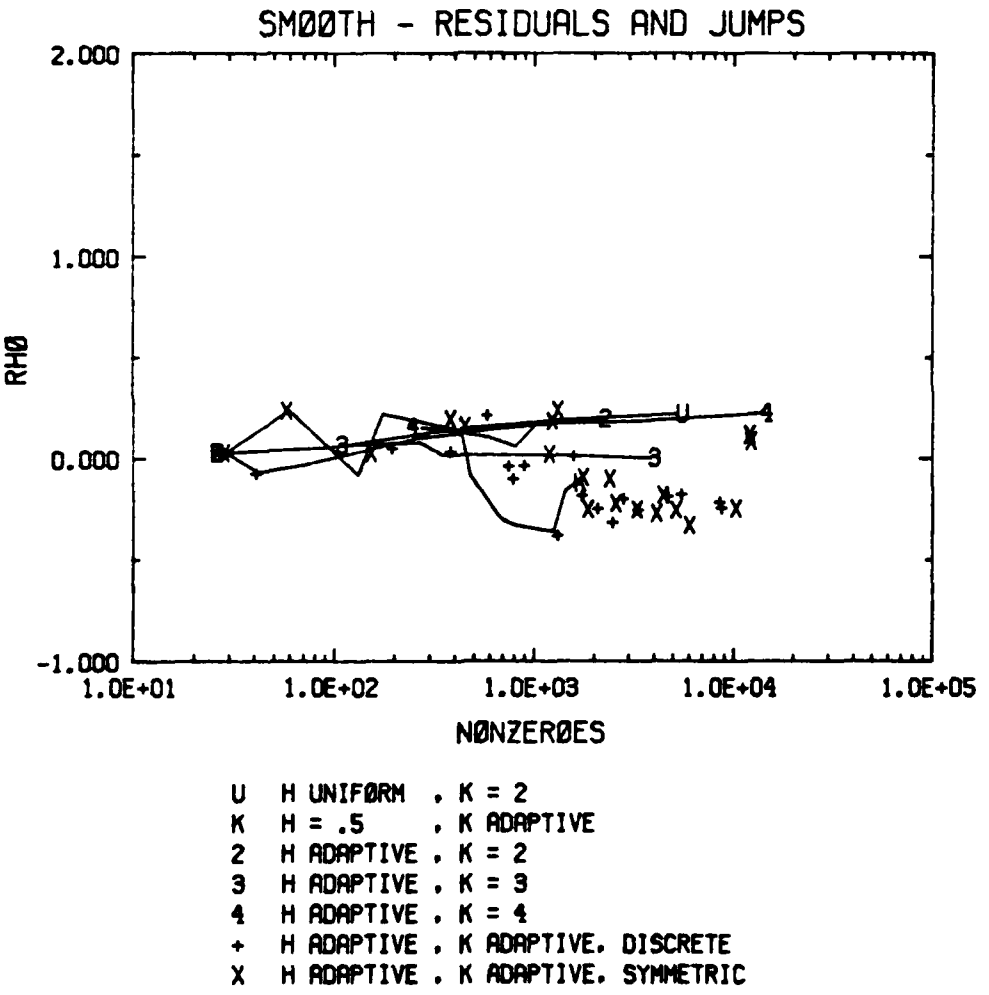


Figure 7-16: Problem 2 - approximate error

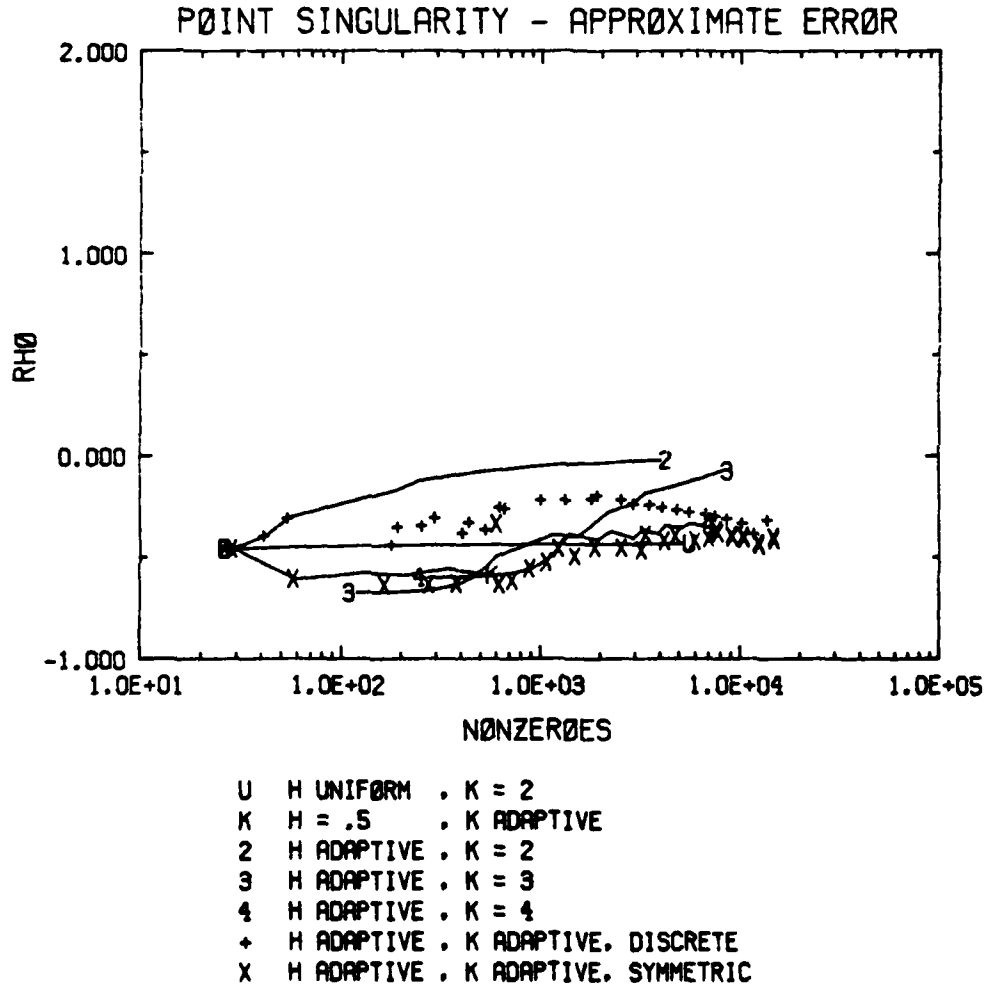


Figure 7-17: Problem 2 - residuals and jumps

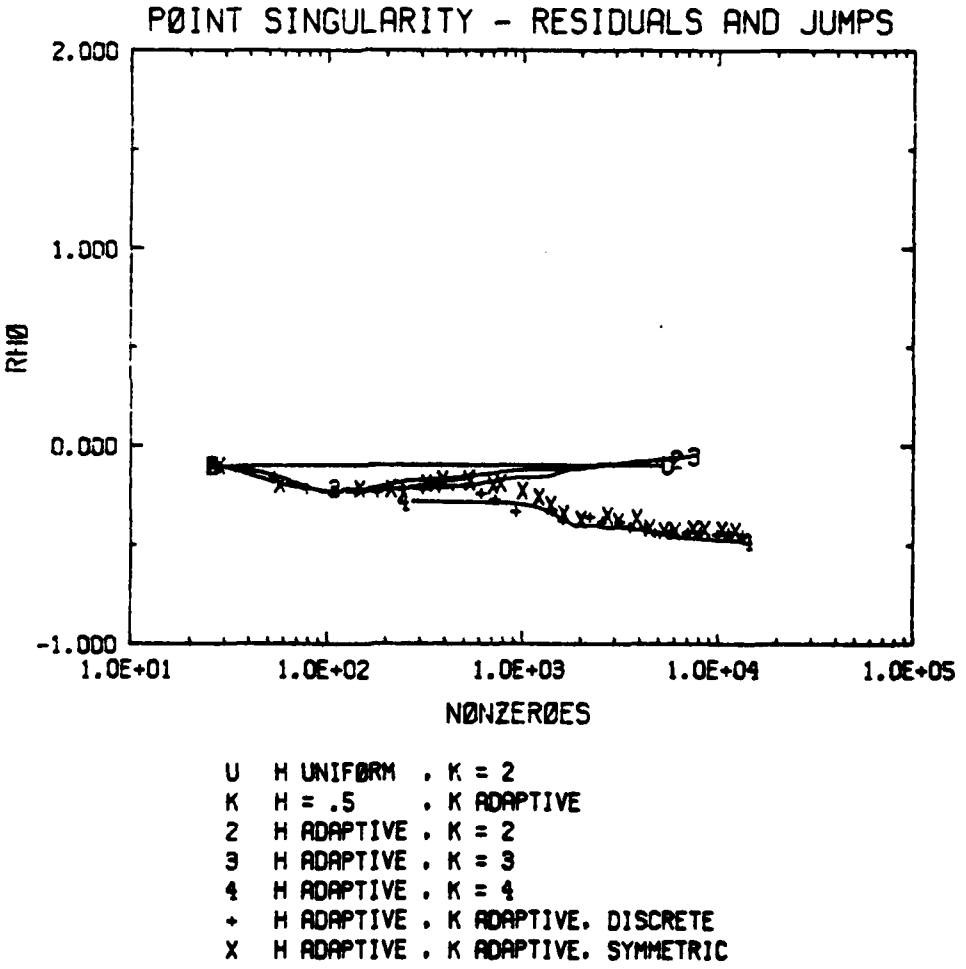


Figure 7-18: Problem 3 - approximate error

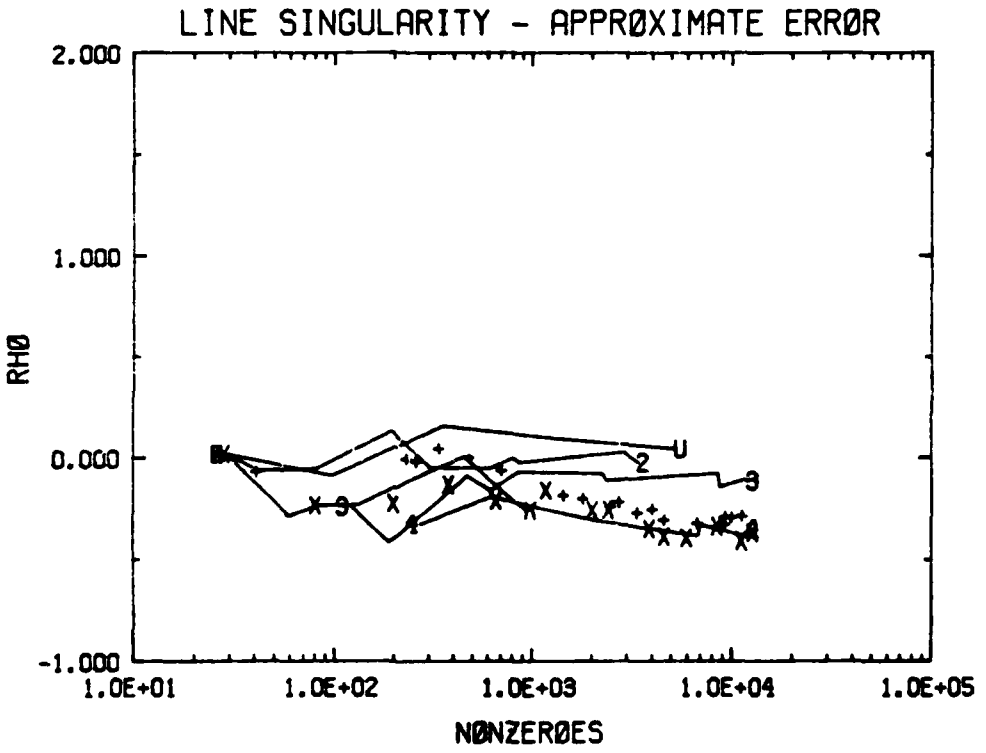
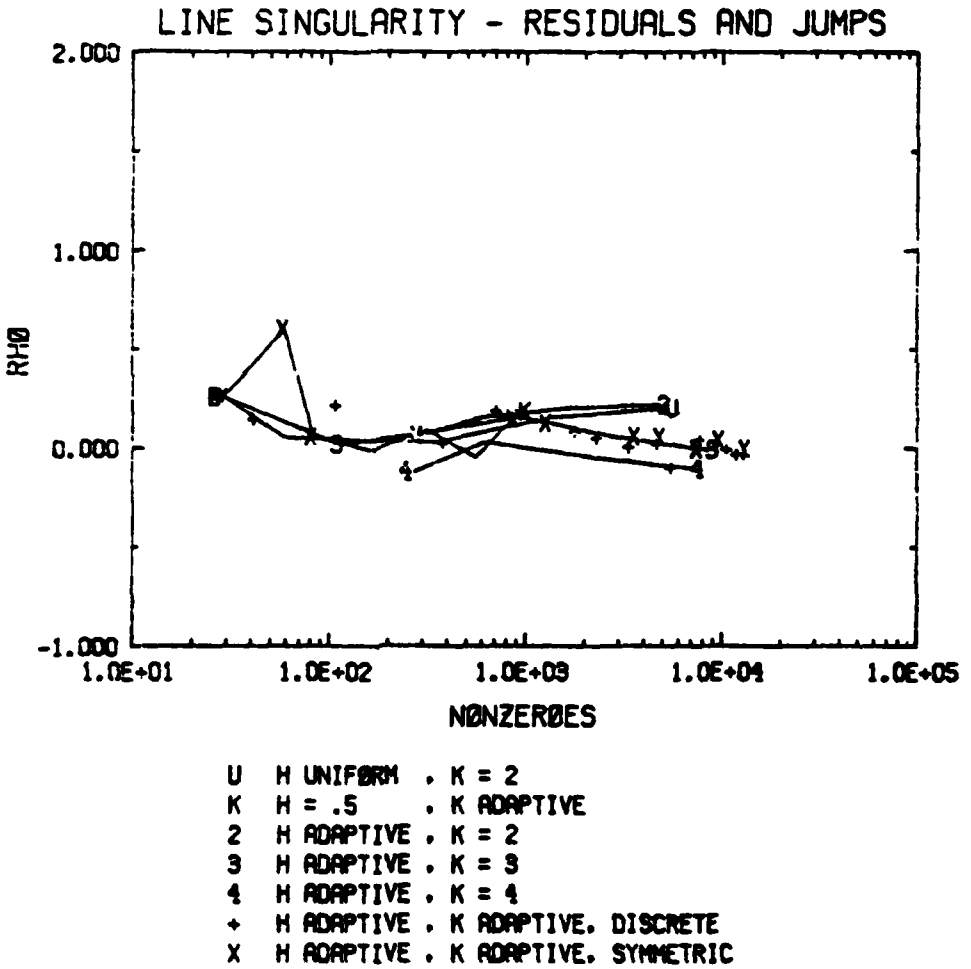


Figure 7-19: Problem 3 - residuals and jumps



Bibliography

[1] S. Agmon.
Lectures on Elliptic Boundary Value Problems.
Van Nostrand, New York, 1965.

[2] I. Babuška and M. R. Dorr.
Error estimates for the combined h and p versions of the finite element method.
Technical Report BN-951, Institute for Physical Science and Technology, University of Maryland, 1980.

[3] I. Babuška, I. N. Katz, and B. A. Szabo.
Hierarchic families for the p-version of the finite element method.
In Proceedings of the Third IMACS International Symposium on Computer Methods for Partial Differential Equations, pages 278-286. Lehigh University, Bethlehem, Pennsylvania, 1979.

[4] I. Babuška and A. Miller.
A posteriori error estimates and adaptive techniques for the finite element method.
Technical Report BN-968, Institute for Physical Science and Technology, University of Maryland, 1981.

[5] I. Babuška and W. C. Rheinboldt.
A posteriori error analysis of finite element solutions for one-dimensional problems.
SIAM Journal on Numerical Analysis 18:565-589, 1981.

[6] I. Babuška and W. C. Rheinboldt.
A posteriori error estimates for the finite element method.
International Journal for Numerical Methods in Engineering 12:1597-1615, 1978.

[7] I. Babuška and W. C. Rheinboldt.
Analysis of optimal finite-element meshes in R^1.
Mathematics of Computation 33:435-463, 1979.

[8] I. Babuška and W. C. Rheinboldt.
Error estimates for adaptive finite element computations.
SIAM Journal on Numerical Analysis 15:736-754, 1978.

[9] I. Babuška and W. C. Rheinboldt.
On the reliability and optimality of the finite element method.
Computers and Structures 10:87-94, 1979.

[10] I. Babuška and W. C. Rheinboldt.
Reliable error estimation and mesh adaptation for the finite element method.
In Computational Methods in Nonlinear Mechanics, pages 67-108. North-Holland, New York, 1980.

[11] I. Babuška, B. A. Szabo, and I. N. Katz.
The p-version of the finite element method.
SIAM Journal on Numerical Analysis 18:515-545, 1981.

[12] I. Babuška and B. A. Szabo.
On the rates of convergence of the finite element method.
Technical Report 80/2, Center for Computational Mechanics, 1980.

[13] R. E. Bank and A. Sherman.
A multi-level iterative method for solving finite element equations.
In Proceedings of the Fifth Symposium on Reservoir Simulation, pages 117-126. Society of Petroleum Engineers of AIME, Dallas, 1979.

[14] R. E. Bank and A. Sherman.
A refinement algorithm and dynamic data structure for finite element meshes.
Technical Report 166, University of Texas Center for Numerical Analysis, 1980.

[15] R. E. Bank and A. Weiser.
In preparation.

[16] A. Brandt.
Multi-level adaptive solutions to boundary-value problems.
Mathematics of Computation 31:333-390, 1977.

[17] P. G. Ciarlet.
The Finite Element Method for Elliptic Problems.
North-Holland, Amsterdam, 1978.

[18] S. C. Eisenstat and M. H. Schultz.
Computational aspects of the finite element method.
In The Mathematical Foundations of the Finite Element Method with Applications to Partial Differential Equations, pages 505-524. Academic Press, College Park, Maryland, 1972.

[19] S. C. Eisenstat.
Private communication.

[20] I. Fried and S. K. Yang.
Best finite elements distribution around a singularity.
AIAA Journal 10:1244-1246, 1972.

[21] D. B. Gannon.
Self Adaptive Methods for Parabolic Partial Differential Equations.
PhD thesis, University of Illinois, 1980.

[22] A. George and J. W. H. Liu.
Computer Solution of Large Sparse Positive Definite Systems.
Prentice-Hall, New Jersey, 1981.

[23] A. George.
Nested dissection of a regular finite element mesh.
SIAM Journal on Numerical Analysis 10:345-363, 1973.

[24] H.-J. Reinhardt.
A posteriori error estimates for the finite element solution of a singularly perturbed linear ordinary differential equation.
SIAM Journal on Numerical Analysis 18:406-430, 1981.

[25] W. C. Rheinboldt and C. K. Mesztenyi.
On a data structure for adaptive finite element mesh refinements.
ACM Transactions on Mathematical Software 6:166-187, 1980.

[26] D. J. Rose and G. F. Whitten.
A recursive analysis of dissection strategies.
In Sparse Matrix Computations, pages 59-83. Academic Press, New York, 1976.

[27] A. Schatz.
An observation concerning Ritz-Galerkin methods with indefinite bilinear forms.
Mathematics of Computation 28:959-962, 1974.

[28] G. Strang and G. J. Fix.
An Analysis of the Finite Element Method.
Prentice-Hall, New Jersey, 1973.

[29] R. L. Taylor.
On completeness of shape functions for finite element analysis.
International Journal for Numerical Methods in Engineering 4:17-22, 1972.

[30] J. R. Van Rosendale.
Rapid Solution of Finite Element Equations on Locally Refined Grids by Multi-Level Methods.
PhD thesis, University of Illinois, 1980.

[31] A. Weiser, S. C. Eisenstat, and M. H. Schultz.
On solving elliptic equations to moderate accuracy.
SIAM Journal on Numerical Analysis 17:908-929, 1980.


DATE
ILME
8